ENHANCING PHOTOMETRIC REDSHIFTS FOR THE ERA OF PRECISION COSMOLOGY

JOHN YUE HAN SOO

Thesis presented for the Degree of
Doctor of Philosophy

Department of Physics and Astronomy
Faculty of Mathematical and Physical Sciences
University College London

September 3, 2018

SUPERVISORS:
Dr Benjamin Joachimi
Prof Ofer Lahav

EXAMINERS:
Dr Filipe Abdalla (University College London)
Prof Matt Jarvis (University of Oxford)

DECLARATION

I, John Yue Han Soo, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

London, September 3, 2018

John Yue Han Soo

ABSTRACT

Photometric redshifts (photo-z's) are vital for the success of current and forthcoming cosmological galaxy surveys. This work focuses on three different approaches to enhance photo-z's. Firstly, we study the extent to which galaxy morphology improves photo-z's. Using artificial neural networks, we compare the performances of several morphological parameters and find that galaxy size and surface brightness bring about the most improvement to photo-z's in bright samples. When multiple morphological parameters are used, the improvement in scatter reaches as high as 12% for the Main Galaxy Sample (MGS) of the Sloan Digital Sky Survey (SDSS). We also find that the improvement becomes significant under suboptimal conditions: when surveys have limited numbers of bands, low quality photometry, or an imperfect star-galaxy separator.

Next we study aspects of photo-z probability density functions (PDFs) and the resulting redshift distributions of galaxy samples in the context of the Canada-France-Hawaii Telescope Stripe-82 (CS82) Survey. We discover that, while galaxy morphology brings marginal improvement to both, we are able to produce accurate redshift distributions using a single photometric band and multiple galaxy morphological parameters, and apply this to the CS82 survey. As part of the photo-z Working Group of the Large Synoptic Survey Telescope Dark Energy Science Collaboration (LSST-DESC), we use several metrics to assess the performances of two state-of-the-art photo-z codes, ANNZ2 and DELIGHT, and concluded that the photo-z's produced by both are close to the standard for the current photo-z requirements of LSST.

Finally, we explore the performances of multiple photo-z codes on narrowband surveys. Using simulated and real data from the 40-narrowband Physics of the Accelerating Universe (PAU) Survey, we find that the hybrid spectral template-machine learning code DELIGHT outperforms monolithic machine learning as well as template codes. Using the large suite of spectral templates and well-calibrated additional broadband fluxes, we are able to produce competitive photo-z's close to the nominal PAU requirement at 40% quality cut.

We believe these method would be useful for the next generation of photometric surveys, like Euclid and LSST.

IMPACT STATEMENT

PROBLEM STATEMENT

The advancement of cosmology research has experienced a breakthrough in recent years with the availability of high resolution images of the sky, made possible by the use of large advanced telescopes and international exchange of expertise. In this era of precision cosmology, astrophysicists continue to explore deeper into the dark universe, and *redshifts*, the third dimension to the sky map, play an important role in characterising the dynamics and structure of our Universe. However, the process of obtaining redshifts spectroscopically is time consuming and expensive.

With the availability of large photometric datasets, various methods of obtaining redshifts photometrically have been developed, and these redshifts are known as *photometric redshifts*. Despite not being as accurate as spectroscopic redshifts, they are widely sought after by many cosmological researches such as weak lensing, galaxy clustering and large-scale structures. Therefore it is the interest of many to improve the current quality of photometric redshifts to characterise our universe in a more accurate and precise way.

RESEARCH WORK

In this work, we used galaxy morphology (e.g. surface brightness, angular radius, ellipticity etc.) to improve photometric redshifts. We conducted a comprehensive study on the impact of galaxy morphology on photometric redshifts using 4 different photometric redshift algorithms on galaxy data from 10 different sky surveys and 3 simulated catalogues of varying depths in magnitudes. We explored and used various performance metrics to characterise the point estimates, probability distribution functions, and galaxy redshift distributions of photometric redshifts on samples which are representative of current and future sky surveys. In the process, we calibrated and optimised photometric redshift algorithms in the context of both broadband and narrowband surveys.

IMPACT

This work has a direct impact to the photometric redshift research community, it adds new knowledge on the efforts to improve the existing methodology. Some of the photometric redshift algorithms used in this work were originally developed by researchers at UCL, and the use and reliability of these algorithms have been further promoted among the astrophysics community. In the process, the history and current development of photometric redshifts have been studied and evaluated, and this knowledge adds value to the astrophysics curriculum to equip future researchers.

Collaborative work was established with researchers from the Canada-France-Hawaii Telescope Stripe-82 Survey (CS82), Large Synoptic Survey Telescope Dark Energy Science Collaboration (LSST-DESC) and the Physics of the Accelerating Universe (PAU) Survey, thus playing a role in international knowledge transfer. Photometric redshift catalogues have been produced in the

process, these data products enable further cosmology and extragalactic astrophysics researches, thus advancing our understanding of the Cosmos.

The impact of this work also extends to the fields of data science and machine learning, as this work employed machine learning algorithms like artificial neural networks, boosted decision trees, k-nearest neighbours and Gaussian processes on large data sets. Particularly, the knowledge gained associated with non-parametric inferential statistical methods, missing data problems, data representativeness and sample reweighting could be applied to any discipline where data correlation and prediction is needed.

ACKNOWLEDGEMENTS

Three years and eight months flew by really quickly! It was as if yesterday when I first stepped into UCL, nervously climbing up the squeaky staircases of the Kathleen Lonsdale Building to get to Benjamin's office for our first meeting. As if it was in an instance I evolved from a fresh PhD student who had big hopes and dreams yet doesn't know a thing, to a PhD holder who has more realistic hopes and dreams yet still doesn't know many things... I guess this was just the typical life of a PhD student, since the wiser ones before me like to say, "it's the process that matters!"

My four years in the UK were full of ups and downs, happy and depressing times... At the time of writing this I'm probably in deep stress trying to hit the submission deadline, but I have faith that all this will be over sooner than I imagined. I wouldn't have made it through without the support of many important people in the process, therefore I would like to spend a few paragraphs below to acknowledge and appreciate them.

Firstly, I would like to thank various UCL staff for making my PhD in UCL even possible at the first place. I would like to thank **Prof. Serena Viti**, who was the graduate admission tutor three years ago, for writing that informal PhD acceptance letter in time for me to get my visa and application done, fulfilling my dream to pursue my PhD in the UK. I would also like to thank **Prof. Ian Howard** for constantly monitoring my progress and ensuring that I have the correct student status while I am at UCL. I would also like to specially thank **Prof. Phil Chan**, **A. Prof. Paul Lim** and **Dr. Cindy Ng** from the National University of Singapore (NUS) for writing reference letters for my PhD applications.

Next I would like to thank the most important person in this section: **Dr. Benjamin Joachimi** – thank you for accepting me as your PhD student! I especially want to mention the few e-mail exchanges we had prior to my admission, where you had been really kind to me even before deciding to become my PhD supervisor. I am truly grateful to have the best supervisor in the world, and I really am not exaggerating! Thank you Benjamin for always understanding me, being patient with me, tolerating with me and also leading me through the learning process. Besides, you have been a very good role model to me: always on time, a family man, very well organised, and always calm at all situations. You are the kind of person I want to be if I ever become a university lecturer in the future. I would also like to thank **Prof. Ofer Lahav** for being my secondary supervisor, for introducing me to the LSST-DESC PZ Working Group, allowing me to apply the knowledge I have on it, and getting to know many photo-z experts in the field. I'm always amazed by how you are able to manage so many people, grants and projects at the same time!

In the following bullet points I would like to acknowledge various groups of people from the UCL Astrophysics Department who have been of great help to me throughout the past four years:

- *My fellow PhD homies*, whom together we started our PhDs at the same time. Thank you for all the fun times we had (mostly during the first year when we were freer). Special

shout-out to **Davide Gualdi** for being my Chinese food buddy and one of my best friends in the department, and also **Thomas Wilson** for being the friendliest person I've ever met!

- *My PhD seniors*, for giving me countless both research related and life lessons. Special shout-out to **Daniela Saadeh**, **Robert Schuhmann** and **Teresita Suarez Noguez**, I really enjoy chatting with all of you, thank you so much! Special thanks to **Antonella Palmese** too, despite always having busy schedules, you helped me out quite a bit with the photo-z code ANNZ2. Not to forget **Gioacchino Accurso**, who constantly organizes Marvel movie nights for all of us!

- *My postdoc partners* who have helped me significantly in my research. I would like to specially thank **Bruno Moraes** for guiding me a lot throughout my 'morpho-z' paper (like, really a lot...); **Will Hartley** for helping me with BPZ and the reweighting algorithm; **Iftach Sadeh** for helping me with ANNZ2; and **Boris Leistedt** for helping me with DELIGHT, and for giving good advice.

- *'Neighbours' in my office*: **Harry Johnston**, **Constance Mahony**, **Nicolas Angelides**, **Jonathan Braden** and **Susan Pyne**. Thank you for bearing with me as I tend to breath really loudly and stay up really late at night...

- The *computer experts* of the department: **John Deacon**, **Edward Edmondson** and also **Sree Balan**, thank you for replying my e-mails and helping me setup things that I could have easily done. I think I know Linux better by now!

- Our *astrophysics administrators*, **Kay Nakum** and **Anita Maguire** for being really helpful to me!

- *UCLO Staff*, specially **Francisco Diego**, **Nicolas Clarke**, **Simon Steel** and **Sarah Hutton**, I really enjoyed teaching at the observatory and managing the annual astrophysics festival 'Your Universe' with you all. Special shout-out to **Claudio Arena** too, it was great working with you too!

Moving on from UCL, I would like to offer my gratitude to the **Ministry of Education Malaysia**, thank you for sponsoring me through the MyBrainSc scholarship. I wouldn't even have dreamt of coming to the UK for my PhD if it wasn't for the generous funding given! As part of my gratitude towards the Malaysian government, I will dutifully hold on to my promise, return and serve the people of Malaysia with the skills and knowledge I acquired.

I would also like to thank some people I meet at the International Lutheran Student Centre (ILSC), where I spent three years living at. Special thanks to the administrators for making my stay there a memorable one, a special shout-out to **Karen Nystrom**, thank you for booking the Club Room for me every week! Also thank you to pastors **Oliver Matri** and **Sarah Farrow** for supporting my student cell group that we hold weekly! Special shout-out my fellow flatmates too, thank you for tolerating my loud piano playing and door-banging when I return really late at night.

Next, I would like thank a special group of people who looked after me spiritually: the congregation of Elim Full Gospel Chinese Church (EFGCC). EFGCC is one of the best churches I have ever attended, and through that I have made lifelong friendships that I would never exchange

for anything else. Thank you for accepting me as part of your big family. I wouldn't be able to list everyone here, but I would like to specially mention **Pastor Aaron Chan**, **Anna Chan**, **Kit Moi Wong** and **Yin Chau Kan**, thank you for being great spiritual mentors to me. I also want specially acknowledge a few friends who have been great encouragement to me throughout my crazy study schedule: **Alicia Chung**, **Cindy Chan**, **Diandian Deng**, **Renjie Yun**, **Eric Chen**, **Hanting Wang**, **Judge Zuo**, **Aunty Judy Candlish**, **Michelle Zhang**, **Stephanie Zhang**, **Aunty Sue Lo**, **Sunnie Mu**, **Wendy Wang** and **Wesley Gilead**. Thank you all for being there for me!

Finally, I would like to thank my family for their continuous encouragement and support for me to pursue my PhD. I want to thank both my aunts living in the UK, **Auntie Catherine Derrick** and **Auntie Lynette Amieux**, for being hospitable to me whenever I visited, I really enjoy spending time with you. I want to thank my dad, mom, brother and sister for always keeping me in their prayers, and having so much faith in me, thank you for not giving up on me, and I am honoured to not have let you down. Last but not least, I thank my beloved girlfriend **Cathryn Chang**: you are the most important person in my life, thank you for being by my side through the many ups and downs. This PhD is dedicated to you, I love you, and cheers to a better future with you!

Ultimately, I thank God for making all things possible, for creating all the miracles in the process, allowing me to pursue my dreams. All glory to God, *Gloria in excelsis Deo*!

John Yue Han Soo
UCL, United Kingdom
30 June 2018

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

2dF	2-degree Field
2MASS	2 Micron All Sky Survey
2SLAQ	2dF-SDSS LRG And QSO
AAO	Anglo-Australian Observatory
AAT	Anglo-Australian Telescope
AGES	AGN Galaxy Evolution Survey
AGN	Active Galactic Nuclei
ALHAMBRA	Advance Large Homogeneous Area Medium-Band Redshift Astronomical
ANN	Artificial Neural Network
ANS	Astronomical Netherlands Satellite
APO	Apache Point Observatory
BAO	Baryon Acoustic Oscillations
BASS	Beijing-Arizona Sky Survey
BDT	Boosted Decision Tree
BFGS	Broyden–Fletcher–Goldfarb–Shanno
BOSS	Baryon Oscillation Spectroscopic Survey
CANDELS	Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey
CAS	Catalogue Archive Server
CCD	Charge-Coupled Devices
CDF	Cumulative Distribution Function
CfA	Centre for Astrophysics
CFHT	Canada-France-Hawaii Telescope
CFHTLS	CFHT Legacy Survey
CFIS	Canada-France Imaging Survey
CMASS	Constant Mass
CMB	Cosmic Microwave Background
CNOC2	Canadian Network for Observational Cosmology 2
COSMOS	Cosmic Evolution Survey
CRPS	Continuous Ranked Probability Score
CS82	CFHT Stripe-82
CSS-OS	Chinese Space Station Optical Survey
CWW	Coleman, Wu and Weedman
CZR	Colour-Redshift Relation
DECaLS	Dark Energy Camera Legacy Survey
DEIMOS	Deep Imaging Multi-Object Spectrograph
DES	Dark Energy Survey
DESC	Dark Energy Science Collaboration
ELG	Emission Line Galaxy

ESO	European Southern Observatory
FIRST	Faint Images of the Radio Sky at Twenty centimetres
FLRW	Friedmann-Lemaître-Robertson-Walker
FWHM	Full Width at Half Maximum
GALEX	Galaxy Evolution Explorer
GAMA	Galaxy And Mass Assembly
GOODS	Great Observatories Origins Deep Survey
GP	Gaussian Process
HDF	Hubble Deep Field
HR	Hertzsprung-Russell
HSC	Hyper Suprime-Cam
HST	Hubble Space Telescope
HUDF	Hubble Ultra Deep Field
IAU	International Astronomical Union
IFU	Integral Field Unit
IMACS	Inamori Magellan Areal Camera and Spectrograph
INT WAS	Isaac Newton Telescope Wide-Angle Survey
IQR	Interquartile Range
JPL	Jet Propulsion Laboratory
IUE	International Ultraviolet Explorer
KiDS	Kilo-Degree Survey
KNN	K-Nearest Neighbour
KS	Kolmogorov-Smirnov
LAS	Large Area Survey
LBQS	Large Bright Quasar Survey
ΛCDM	Λ Cold Dark Matter
LIGO	Laser Interferometer Gravitational-wave Observatory
LMC	Large Magellanic Cloud
LOWZ	Low Redshift
LRG	Luminous Red Galaxy
LSST	Large Synoptic Survey Telescope
MCMC	Markov Chain Monte Carlo
MCP	Megamaser Cosmology Project
MICE	Marenostrum Institut de Ciències de l'Espai
MGS	Main Galaxy Sample
MLP	Multi-Layer Perceptron
MOND	Modified Newtonian Dynamics
MOS/SIS	Multi-Object Spectrograph / Subarcsecond Imaging Spectrograph
NOAO	National Optical Astronomy Observatory
NRAO	National Radio Astronomy Observatory
NYU-VAGC	New York University Value-Added Galaxy Catalogue
ORM	Observatorio del Roque de los Muchachos
PAU	Physics of the Accelerating Universe

PCA	Principal Component Analysis
PRIMUS	Prism Multi-object Survey
PDF	Probability Density Function
PHAT	Photometric redshift Accuracy Testing
PIT	Probability Integral Transform
PSF	Point Spread Function
PZ	Photometric Redshift
QSO	Quasi-Stellar Object
RCS	Red-sequence Cluster Survey
RCSLenS	Red Cluster Sequence Lensing Survey
redMaGiC	red-sequence Matched filter Galaxy Catalogue
S82-MGC	Stripe-82 Massive Galaxy Catalogue
SCP	Supernova Cosmology Project
SDSS	Sloan Digital Sky Survey
SNIa	Type Ia Supernova
SED	Spectral Energy Distribution
SPLASH	Spitzer Large Area Survey with HSC
SSP	Simple Stellar Population
SVM	Support Vector Machine
TDI	Time-Delay-and-Integrate
TeVeS	Tensor-Vector-Scalar Gravity
TMVA	Toolkit for Multivariate Data Analysis
UKIDSS	UKIRT Infrared Deep Sky Survey
UKIRT	United Kingdom Infrared Telescope
VIKING	VISTA Kilo-degree Infrared Galaxy
VIMOS	Visible Multi-Object Spectrograph
VIPERS	VIMOS Public Extragalactic Redshift Survey
VISTA	Visible and Infrared Survey Telescope for Astronomy
VLBI	Very Long Baseline Interferometry
VLT	Very Large Telescope
WFIRST	Wide Field Infrared Survey Telescope
VVDS	VIMOS VLT Deep Survey
WHT	William Herschel Telescope
WISE	Wide-field Infrared Survey Explorer
XMDS	X-ray multi-mirror mission Medium Deep Survey

Part I

INTRODUCTION

0

ROADMAP

The fear of the Lord is the beginning of wisdom, and the knowledge of the Holy One is understanding.

Proverbs 9:10

Humans have always been curious about what lies beyond the Earth we live in. Ever since the civilisation of mankind, humans have been putting efforts into the creation of devices which could help us understand the Cosmos better. Humans invented the telescope to get a better look at the celestial bodies in the sky, and as human's knowledge of science advanced, they started inventing more sophisticated equipment to study the properties of these objects. Take the moon for example, in the past people came up with crazy stories and theories about what it is made of, and why it is there. Fast forwarding to today, we now have a good idea of what the moon is, and humans have even set foot on it.

When observing celestial bodies, a common question one might ask is: "will we be able to get there?" That's when distance measurements come in: humans became interested in knowing how far these celestial objects are away from us, and how long it takes us to get to them. With the astrophysics knowledge we have today, I am sure that everyone reading this knows that a casual tour to a planet or a galaxy is not feasible at the moment, but our motivation for cosmological distance measurements has changed! We started off measuring distances so that we could calculate the time to arrive the destination, but now, we measure distances in order to understand our universe better, discover its past, and predict its future. Stepping into the 21st century, we also see another shift in methodology: from theoretically-driven approaches, we now dive into data-driven approaches. Facing datasets with unprecedented sizes, we find ourselves wrestling with new statistical challenges for precision and accuracy. In this era we find ourselves bridging astrophysics with machine learning and data science.

This thesis is about improving photometric redshifts (or photo-z's in short), which if looked from the bigger picture, is actually just improving distance measurements to the furthest objects in our universe. Photo-z research has been around for decades, and with the emergence of deeper and bigger sky surveys in the 21st century, we expect to see it continue to advance. The improvement of photo-z measurements may not be a 'ground-breaking' discovery leading to a Nobel prize, yet it is essential as current cosmology research depend on large reliable and accurate photo-z catalogues to progress, to help us understand the universe that we live in.

This thesis is written as follows. In Chapter 1 we first discuss the big picture by defining the physics and cosmology of this work. We use the distance ladder narrative to introduce the reader to distance measurements: distances spanning from the nearby planets, to the Baryon Acoustic

Oscillations (BAO) measured from when the Universe was only about 379 000 years old. We will define the different cosmological distance measurements, and particularly how redshifts are defined using general relativity. In Chapter 2 we proceed to discuss galaxy photometry, morphology and spectroscopy. We will explain how fluxes are measured using the different photometric systems and magnitude fits; we will briefly discuss and define the galaxy morphological parameters used in this study; and we will discuss the physics behind spectroscopy and redshifts.

With all the theories of cosmology, photometry and spectroscopy described, we proceed to discuss the main topic of our work: photometric redshifts. In Chapter 3 we discuss photo-z's in full detail, giving a brief history of it and the basics of its 3 different methods: template fitting, empirical methods, and machine learning methods. In this chapter we will also discuss the various recent efforts to improve photo-z estimations, the current photo-z requirements by various large surveys, which leads to the point where we state the motivation of this thesis.

In the following 2 chapters we describe the methodologies and sources of data used in this work. In Chapter 4 we list down all the photometric surveys, simulated galaxy catalogues and redshift surveys whose galaxy photometry, morphology and redshift data were used in this research. Each survey or catalogue will be carefully described, any sample selection cuts used on these samples will also be discussed. In Chapter 5 we introduce the 4 different photo-z algorithms used in this paper: ANNZ, ANNZ2, BPZ and DELIGHT, where we describe their algorithms and the optimisation used. We also briefly introduce the spectral energy distribution (SED) templates used in this work, and the metrics used to evaluate our photo-z performances.

The following 3 chapters are the main results of this thesis. Chapter 6 discusses the impact of galaxy morphological parameters on photo-z's using the machine learning codes ANNZ and ANNZ2, through 4 different samples of different depths, namely the Main Galaxy Sample (MGS), Luminous Red Galaxies (LRGs), the Stripe-82 Sample and the CFHT Stripe-82 (CS82) Sample. This is a comprehensive study on how morphology makes an impact on photo-z's in different settings, e.g. in elliptical and spiral galaxy samples, between bright and faint samples, in samples reweighted to a deep photometric survey, and in surveys with suboptimal conditions.

Chapter 7 moves away from point estimates and focuses on the study of photo-z probability density functions (PDFs) and redshift distributions of galaxy samples. The beginning of the chapter studies the impact of galaxy morphology on photo-z PDFs and redshift distributions, and the results obtained are applied to produce a photo-z catalogue for the CS82 Survey. The later part of this chapter is part of a collaborative work with the Large Synoptic Survey Telescope Dark Energy Science Collaboration Photometric Redshift (LSST-DESC PZ) working group, where different metrics were explored to characterise photo-z PDFs and redshift distributions obtained in the context of LSST. Particularly, we will compare the performance of 2 state-of-the-art photo-z algorithms: ANNZ2 and DELIGHT. We evaluate their photo-z products, as well as review the metrics used to evaluate their performances.

Finally, Chapter 8 discusses photo-z's in the context of a 40-narrowband survey, specifically the Physics of the Accelerating Universe (PAU) Survey. Using ANNZ, ANNZ2, BPZ and DELIGHT, we compare the performances of each of these algorithms on both simulated and real galaxy data. We put a particular focus on DELIGHT on this chapter, where we use the simulated and real galaxy data to optimise the settings and hyperparameters of DELIGHT, particularly on the choice of templates, and methods used to calibrate and match fluxes from different surveys.

The thesis is concluded in Chapter 9, and the possibilities of future work and outlook are also briefly discussed.

1

COSMOLOGY

Then God said, "Let there be light"; and there was light.

Genesis 1:3

1.1 INTRODUCTION

1.1.1 *General Relativity and the FLRW Metric*

Cosmology is the study of the dynamics of large structures in our Universe, it is concerned with the fundamental questions of the formation, evolution and ultimate fate of our Universe. We start by assuming the *cosmological principle*, which states that the distribution of matter-energy in the universe on a large scale ($\geq$ 100 Mpc) is homogeneous and isotropic, where homogeneous means that it is 'the same everywhere', while isotropic means that it is 'the same when looked at every direction'.

A more quantitative and theoretical description of our Universe is provided by *general relativity*, which tells us that the gravitational field and force we know is generated from the geometry of spacetime itself. This means that all information on how gravity will act on a mass particle held in a particular condition will be contained in a *metric tensor*. The simplest form of a metric tensor for a universe without matter and radiation is called *de Sitter metric* (de Sitter 1911):

$$ds^2 = -c^2\,dt^2 + dx^2 = -c^2\,dt^2 + dr^2 + r^2\,d\theta^2 + r^2\,\sin^2\theta\,d\phi^2, \tag{1}$$

where $-c\,dt$ is the temporal contribution with c the speed of light, and dx the spatial contribution, with $\{r, \theta, \phi\}$ representing the radial distance, polar angle, and azimuthal angle of spherical coordinates. It can be seen as a set of coordinates for the 4-dimensional Euclidean spacetime. However, such a spacetime does not explain our Universe, the more appropriate metric which describes a homogeneous, isotropic expanding or contracting universe is known as the *Friedman-Lemaître-Robertson-Walker* (FLRW) metric (Friedman 1922; Lemaître 1931; Robertson 1935; Walker 1937):

$$ds^2 = -c^2\,dt^2 + a(t)^2\left[\frac{dr^2}{1 - k\left(\frac{r}{r_0}\right)^2} + r^2\,d\theta^2 + r^2\,\sin^2\theta\,d\phi^2\right]. \tag{2}$$

$a(t)$ is the time dependent scale factor, an expanding universe has a scale factor $\frac{d}{dt}a(t) = \dot{a}(t) > 0$, while an accelerating universe has $\frac{d^2}{dt^2}a(t) = \ddot{a}(t) > 0$. Our Universe has been shown to be an expanding universe by Hubble (1929a), and an accelerating universe by Riess et al. (1998) and the Supernova Cosmology Project (SCP, Perlmutter et al. 1999), therefore for the practical purpose of this work, we will assume an accelerating expanding universe unless stated otherwise. The present value of a, denoted $a(0)$ or a_0 is set to 1.

The FLRW metric admits 3 kinds of spatial curvature: closed (spherical, $k = 1$), flat ($k = 0$) and open (hyperbolic, $k = -1$). The flatness of our Universe can be observed from the measurements of the cosmic microwave background (CMB) radiation, in which recent results from Planck (Ade et al. 2016) showed that our Universe appears to be spatially flat to a 1σ accuracy of 0.25%. For the purpose of this work, we will assume a flat universe (theoretically motivated by inflation), setting $k = 0$ and Equation 2 becomes

$$ds^2 = -c^2\,dt^2 + a^2(t)\left[dr^2 + r^2\,d\theta^2 + r^2\sin^2\theta\,d\phi^2\right]. \tag{3}$$

As mentioned earlier, general relativity describes the fundamental interaction of gravitation as a result of spacetime being curved by mass and energy. This can be written mathematically, known as the *Einstein field equations* (Einstein 1916):

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \frac{8\pi G}{c^4}T_{\mu\nu}, \tag{4}$$

where $R_{\mu\nu}$ is the Ricci tensor (controls the metric volume growth rate of the metric in the spacetime), $g_{\mu\nu}$ the metric tensor, R the Ricci scalar (compares the deviation of volume between a Euclidean space and the said spacetime) and G the gravitational constant, $G = 6.67 \times 10^{-11}$ m^3 kg^{-1} s^{-2}. $T_{\mu\nu}$ is the energy-momentum tensor, it contains information of all the constituents in our Universe, which includes radiation, baryonic matter and non-baryonic matter. More precisely, T_{00} is the energy density $-\rho c^2$, $T_{i0} = T_{0i}$ the momentum density (or energy flux) and T_{ij} is the stress tensor, in which for the case of a homogeneous and isotropic universe, will have the diagonals of the matrix being the pressure of the substance P. Einstein's field equation tells us that the existence of energy and matter would bend the spacetime around it, causing the Newtonian 'gravitational force' as we know when two massive bodies are close to one another.

The ratio between P and ρ is known as the *equation of state*, $w = \frac{P}{\rho c^2}$. Dust (or slow moving matter in general) has $w = 0$ (since $P \ll \rho c^2$), while relativistic matter (or radiation) has $w = \frac{1}{3}$ (since radiation pressure is $\frac{1}{3}$ of its density due to the randomisation of photos in all directions). We could in fact include a 'curvature density' if we left k arbitrary, which would give us $w = -\frac{1}{3}$, but the discussion on this is beyond the scope of this work.

1.1.2 *The LCDM Cosmological Model*

In assuming a homogeneous, isotropic, flat and accelerating universe, we opted to use the widely accepted Λ *Cold Dark Matter* (ΛCDM) cosmological model for this research. The ΛCDM model suggests the existence of *cold dark matter*, a slow-moving substance which interacts gravitationally but not electromagnetically; and *dark energy*, a substance which causes the acceleration of the universe. At present, dark matter particles have not been detected, however there is indirect

evidence supporting its existence, e.g. the velocity dispersion of galaxies in clusters are higher than expected from their visible mass content (Kapteyn 1922; Zwicky 1933); differential rotation velocity of spiral galaxies not matching the rotational curves expected (Rubin et al. 1980); the baryons accounting only a small fraction of mass of galaxies and clusters based on X-rays emitted (Vikhlinin et al. 2006); and perhaps the more famous evidence – the significant offset between luminous and gravitating components of the Bullet Cluster (Markevitch et al. 2004). For the interest of the reader, modified gravity models which are able to explain these phenomena without the use of dark matter exist, like Modified Newtonian Dynamics (MOND, Milgrom 1983) and its relativistic counterpart, Tensor-Vector-Scalar Gravity (TeVeS, Bekenstein 2004), however these models shall not be discussed further in this work.

In the ΛCDM model, dark energy can be characterised by the cosmological constant Λ, which was a term originally in the Einstein's field equation to explain a static universe but removed later on. With Λ, Equation 4 now becomes

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + g_{\mu\nu}\Lambda = \frac{8\pi G}{c^4}T_{\mu\nu}. \tag{5}$$

Λ being proportional to an energy tensor means that dark energy has energy and momentum. With the subscript Λ denoting dark energy, it can be shown that

$$\rho_\Lambda c^2 = \frac{\Lambda c^4}{8\pi G} = -P_\Lambda, \tag{6}$$

which gives us $w = -1$, making dark energy asserting negative pressure to its surroundings.

1.1.3 *The Friedman Equations*

In order to study theoretically the evolution of substance abundance in our Universe, we insert the FLRW metric (Equation 2) into the Einstein field equation (Equation 4), and get the *Friedman equations*:

$$2\frac{\ddot{a}}{a} + \frac{\dot{a}^2}{a^2} = -\frac{8\pi G}{c^2}P, \tag{7}$$

$$3\frac{\dot{a}^2}{a^2} = 8\pi G\rho. \tag{8}$$

The Friedman equations govern the expansion of space, we can use this equation to yield the time evolution and geometry of the universe as a function of fluid density and pressure. Assuming single fluid-dominated universes, we use the properties of conservation of energy and momentum on the Einstein field equation (i.e. the covariant derivative of the energy-momentum tensor $\nabla_\mu T^{\mu\nu} = 0$) and Equation 8 to produce the *fluid equation*:

$$\frac{1}{\rho}\frac{d\rho}{dt} + \frac{3}{a}\frac{da}{dt}(1+\omega) = 0. \tag{9}$$

We can solve the equation above for a matter-dominated universe ($\rho = \rho_m$ and $w = 0$), radiation-dominated universe ($\rho = \rho_r$ and $w = \frac{1}{3}$) and a dark energy-dominated universe ($\rho = \rho_\Lambda$ and $w = -1$), giving us equations of densities with respect to time:

$$\rho_m = \rho_{m,0} a^{-3}, \qquad \rho_r = \rho_{r,0} a^{-4}, \qquad \rho_\Lambda = \rho_{\Lambda,0}, \tag{10}$$

where the subscript 0 represents the densities of those fluids at the present day. Since we know that a increases with time, from this we can deduce that at early times the density of radiation dominates the universe, however at late time, the the density of radiation decreases much more than matter which results in a matter-dominated universe. The density of dark energy remains constant throughout time, therefore as matter and radiation density continue to decline, we enter an era of a dark energy-dominated universe today.

To study the expansion of the universe, we can solve Equation 7 to get the following:

$$\frac{\ddot{a}}{a} = -4\pi G\rho \left(w + \frac{1}{3}\right). \tag{11}$$

For the universe to accelerate ($\ddot{a} > 0$), we see that the equation above must satisfy $w < -\frac{1}{3}$, in which from the previous paragraph, we know that a universe which is only matter or radiation-dominated would not be able to satisfy this condition. In other words, a universe with a mixture of dark energy and matter or radiation is required for an accelerating universe. Here we assume that w is independent of $a(t)$, there are time dependent models like quintessence (Ratra and Peebles 1988) which is beyond the discussion of this work.

Next we define the *Hubble parameter*, $H = \frac{\dot{a}}{a}$, which defines the the expansion rate of the universe. Assuming a dark energy and matter-dominated universe, using Equation 6 and Equation 8, we get

$$H^2 = \frac{8\pi G}{3}(\rho_m + \rho_\Lambda) = H^2(\Omega_m + \Omega_\Lambda) = H_0^2(\Omega_{m,0} a^{-3} + \Omega_{\Lambda,0}), \tag{12}$$

where H_0 is the Hubble parameter at the present age, also known as the *Hubble constant* in which Hubble (1929a) formulated the famous Hubble equation $v = H_0 d$, where v is the recessional velocity of the galaxy and d the proper distance between the galaxy and the observer. As seen from the equation, the Hubble parameter decreases as time passes. The recent Planck results have shown best fit values of $H_0 = 67.8 \pm 0.9$ km s^{-1} Mpc^{-1} (Ade et al. 2016), however, currently there exists is a tension in the determination of its value. The value of H_0 determined in the nearby Universe is in tension with the value derived based on measurements of the CMB (Freedman 2017). This implies the possibilities of new physics or uncertainties yet to be discovered.

The Ω_i terms in Equation 12 are relative densities, they are defined as

$$\Omega_i = \frac{8\pi G}{3H^2}\rho_i. \tag{13}$$

The terms can also be defined as $\Omega_i = \frac{\rho_i}{\rho_c}$, with respect to their critical densities $\rho_c = \frac{3H^2}{8\pi G}$. In a flat universe, the sum of the relative densities should add up to 1 ($\sum_i \Omega_i = 1$). Cosmologists are interested in these numbers since they tell us the relative densities of the fluids in our Universe at the present time, based on various cosmological probes. The recent measurements of Planck (Ade

et al. 2016) gave the following results for relative densities: $\Omega_{m,0} = 0.31 \pm 0.01$, $\Omega_{\Lambda,0} = 0.69 \pm 0.01$, $\Omega_{r,0} = 0.000093 \pm 0.000005$ and $\Omega_{k,0} = 0.000 \pm 0.005$, which shows that we live in a flat universe made up of 69% dark energy and 31% matter. Ω_m is a sum of the baryon density Ω_b and cold dark matter density Ω_c, which separately yield $\Omega_b = 0.048 \pm 0.001$ and $\Omega_c = 0.258 \pm 0.008$. This shows that cold dark matter is about 5 times as abundant as baryonic matter.

1.2 THE COSMIC DISTANCE LADDER

Since this thesis is about improving photometric redshifts (the empirical distance to galaxies and quasars), it would be helpful to understand where spectroscopic and photometric redshifts lie in the *cosmic distance ladder*. The cosmic distance ladder is the succession of methods by which astronomers determine distances to celestial objects, from the nearest to the furthest. Such a ladder exists since there is no single distance measurement method which could cover all distances, as well as all kinds of objects. In this section we briefly summarise the history of cosmological distance measurements, divided into several sections roughly based on chronology, type of object and distance to that object.

1.2.1 *The Astronomical Unit*

At the base of the distance ladder we start by measuring distances to objects within the Solar System by making *direct measurements*, in which no physical assumptions about nature is needed. We start by defining a distance measurement unit to objects within the Solar System, the *astronomical unit* (AU): the rough distance between the Sun and Earth. The AU used to be defined as the average of Earth's aphelion and perihelion, it is currently precisely defined as 149 597 870.7 km by the International Astronomical Union (IAU, IAU 2012).

The earliest mention of measuring distances to the Sun was done by Aristarchus in 3rd century BCE, in his book *On the Sizes and Distances (of the Sun and Moon)* where he calculated the distance between the Earth and the Sun to be about $380R_\oplus$ (Van Helden 2010), where $R_\oplus$ is the radius of the Earth. It was estimated using trigonometry by measuring the angle between a half Moon and the Sun, and the result was far from correct since the current measured value is about $23000R_\oplus$. Later measurements have been done by others like Hipparchus and Ptolemy, but none have obtained the distance within orders of magnitude correct (Van Helden 2010).

In the 1600s, Kepler's laws of planetary motion were formulated, this aided the first scientifically correct measurement of the distance between the Sun and Earth, which was measured with the help of the *Venus transit* and *trigonometrical parallax* (more on this in Section 1.2.3). Venus transits are very rare events, they are spaced out with a pattern of 105.5, 8, 121.5 and 8 years apart in a 243-year cycle. In 1639, which was also the first scientifically documented Venus transit, Jeremiah Horrocks measured the distance from the Earth to the Sun with an accuracy of up to two thirds of the current distance (Marston 2004), while the next Venus transit in 1769 allowed measurements close to 3% deviation from current value (Smith 1769). The Venus transit remains an important test of astrometry, however as radio telemetry developed, recent Venus transits (like the one in 2004) have been mainly used for extrasolar planet studies (Schneider et al. 2006).

The reader can refer to Teets (2003) for a good review on the mathematics and history of using Venus transits as a distance indicator.

A more accurate measurement of the distance to the Sun was put forth by Simon Newcomb in the 1900s when he made use of the *constant of aberration* κ. Aberration of light is the apparent displacement of light from the celestial object caused by the apparent motion of the object. To correct for aberration, the relative velocity between the observer and light source has to be taken into account. κ was defined as the maximum displacement of a star due to annual aberration (assuming a circular orbit of the Earth), and was approximated to be $\kappa \approx \frac{v}{c}$, the ratio between the Earth's average speed in the Sun's frame and the speed of light, where its accepted value is currently $20.49552''$ (Kovalevsky and Seidelmann 2004). Accounting for aberration in the results from previous Venus transits, Newcomb measured the solar parallax to be $8.80''$ (Newcomb 1906), which is very close to the current value of $8.794143''$. This value was used by the IAU until 1964 (IAU 1964).

1.2.2 *Radar*

Until this point, measurements of distances to celestial bodies within the Solar System had relied on trigonometry and Newton's law of universal gravitation. When radio telemetry was developed in the 1900s, distances to asteroids and planets in the Solar System could be measured using space probes and *radars*, thus the need for parallax calculations for objects within the Solar System has been superseded.

Radar astronomy is a technique to observe nearby celestial objects by reflecting microwaves off them and analysing their reflections. This is based on the simple fact that the distance to an object is the speed of light multiplied with half the time travel of the radar signal, where the constant speed of light was known during this era. Consider a microwave from a ground-based radar which was transmitted and bounced off a celestial object of interest, the power P_r returning to the receiving antenna is given by the *radar equation* $P_\mathrm{r} \propto \frac{1}{d^4}$, where d the distance between the transmitter to the target (full equation in Stimson 1998). From the equation we see that the power received is inversely proportional to the 4th power in distance, therefore the distance between the object and the antenna can be measured when the power of the receiving microwave is measured.

The Jet Propulsion Laboratory (JPL) was the first to use a planetary radar system on Venus, and calculated the AU up to an accuracy of ± 500 km (Malling and Golomb 1961). Later, the Lunar Laser Ranging Experiment of the Apollo mission was set up to detect optical echoes from the moon and measure its distance from Earth (Smullin and Fiocco 1962). Radar data collected in those years were used to measure distances to the inner planets (Ash et al. 1967), and even smaller objects like asteroids and comets (Ostro 1985).

1.2.3 *Stellar Parallax*

Moving on from the Moon, Sun and planets within the Solar System, distance measurements to objects within our Galaxy proved more difficult. At this point a unit much larger than the AU would be more suitable to measure distances to nearby stars and star clusters. One distance that

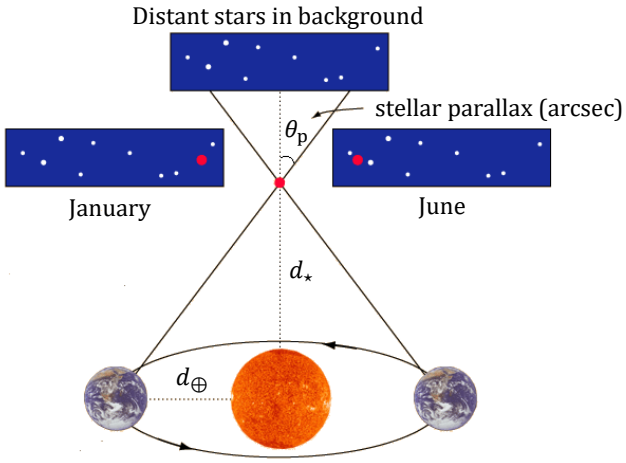

Figure 1: Visualisation of the concept of stellar parallax. Figure obtained and modified from http://hyperphysics.phy-astr.gsu.edu.

may cross the mind is the *light-year* (ly), which is the distance travelled by light in one Julian year (365.25 days), which has a value defined by the IAU (2012) to be 9.4607×10^{15} m (or about 63 241.1 AU). However this unit is seldom seen in astronomy nor cosmology research, instead more commonly used in general science articles and popular culture (and often misunderstood as a unit of time!).

Distances to nearby stars are probably best measured using trigonometric *stellar parallax*. The parallax effect is the apparent shift in position of a relatively nearby object against more distant ones when viewed from different vantage points. The positions of nearby stars are shifted slightly against the background when viewed from the Earth at the two extreme positions of Earth's orbit around the sun (6 months apart, with a distance of 2AU in between). So the parallax angle of a star θ_p is defined to be half of the angular distance that a star appears to move relative to the celestial sphere as Earth orbits the Sun. Using trigonometry,

$$\tan \theta_p = \frac{d_\oplus}{d_\star} \tag{14}$$

where $d_\oplus$ is the distance from the Sun to Earth, and $d_\star$ is the distance from the Sun to the star. This relationship is visualised in Figure 1.

As mentioned in Section 1.2.1, parallax has been used earlier to measure distances to planets, but it wasn't until the 1800s when it could reliably measure distances to stars. Friedrich Bessel was the first to do so, he used a Fraunhofer heliometer at Königsberg Observatory to obtain the distance to the star 61 Cygni (Bessel 1839). It can be measured that most stars visible to the unaided eye in the night sky are within 500 pc from the Sun.

Parallaxes have been measured in arcseconds, until 1900s when the *parsec* (literally 'parallax of one arcsecond', named by Herbert Turner) was introduced (Dyson 1913). A parsec (pc) was defined to be the distance at which 1 AU subtends an angle of $1''$. Generally for stars very far away (small angles), we get the distance to be $d_\star \approx \frac{1}{\theta_p''}$ pc where θ_p is measured in arcseconds (2 pc is $0.5''$, 5pc is $0.2''$ and etc). It is worth noting that parsec is a measure of distance but not angle, in terms of the AU 1 pc$= \frac{648\,000}{\pi} \approx 206\,264.8$ AU, and in terms of light-year 1 pc$= 3.26156$

ly . This however is true only if θ_p is small, for larger angles (or objects much nearer to the Sun), the correct formula for x parsecs should be:

$$d_\star = \frac{1}{\tan\theta_\mathrm{p}}\,\mathrm{AU} = \frac{\pi}{648\,000}\frac{1}{\tan\theta_\mathrm{p}}\,\mathrm{pc} \tag{15}$$

where we see that θ_p [arcsec] $= \frac{\pi}{648\,000}\theta_\mathrm{p}$ [radians] would revert to the previous formula by the approximation of $\tan x \approx x$ for small angles.

Ground based telescopes are limited to parallax angles of about $0.01''$, which means that measurements of stars further than 100 pc would not be accurate due to seeing. The Hipparcos space satellite was launched in the 1990s to measure distances to hundred thousands of stars with a precision of about a milliarcsecond (Perryman et al. 1997), while an even more recent probe is Gaia, launched in 2013, which has an increased precision of 24 microarcseconds, allowing it to measure distances to billions of stars (Prusti et al. 2016). Gaia's first data released allowed parallax measurements of many stars cross-checked with previous probes (Clementini et al. 2017), while its recent second data release (Brown et al. 2018) provided proper motions and parallaxes of 1.3 billion stars.

1.2.4 *The Magnitude System*

Before we move away from trigonometric distance measurements to photometric distance measurements, it is appropriate that we introduce the *magnitude* system at this point. The magnitude system is a logarithmic measure of brightness. There are two kinds of magnitudes used for stars and galaxies: the *apparent magnitude* m is the brightness as it appears in the sky as observed on Earth, while the *absolute magnitude* M_V is the intrinsic brightness of an object as it would appear if placed 10 pc away from the Earth[1]. Since stars are point sources in the sky, the magnitudes are the brightness for that 'point', while magnitudes for galaxies are obtained by using its integrated brightness (more on this in Section 2.1).

Magnitudes are usually measured in a certain passband filter (a band allowing light within a certain wavelength to pass through), and here the subscript V denotes the visual V-band. A magnitude adjusted to take account of radiation across all wavelengths is called a *bolometric magnitude* M_bol, which can be estimated by applying a bolometric correction term to M_V. This correction is needed because hotter objects radiate mostly ultraviolet radiation, while colder objects radiate mostly infrared radiation, as dictated by Planck's law. $M_\mathrm{bol} = 0$ corresponds to a luminosity of 3.0128×10^{28}W (IAU 2015).

As mentioned earlier, the brightness scale of magnitudes is logarithmic, meaning that a change in 1 magnitude corresponds to the change in brightness by 2.51 times (or a change in 5 magnitudes corresponds to the change in brightness by 100 times), mathematically,

$$m_1 - m_2 = -2.5\log_{10}\frac{L_1}{L_2}, \tag{16}$$

where L_i is the luminosity (brightness) of the object. Although Hipparcos and Ptolemy were the first to introduce the magnitude system, it was in fact Norman Pogson who suggested make

1 The definition of M_V for planets and asteroids is different and more complex than that of galaxies and stars, and for the purpose of this work, M_V will assume the definition of the latter.

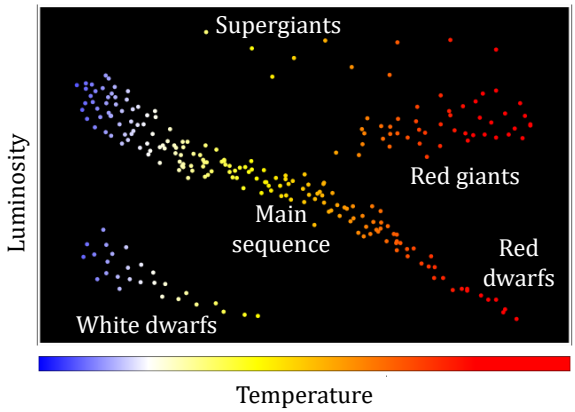

Figure 2: The HR diagram. Figure obtained and modified from http://sunshine.chpc.utah.edu.

this brightness ratio a systematic standard (Pogson 1856), in which later the value $\sqrt[5]{100}$ became known as the *Pogson's ratio*. The logarithmic magnitude system is very helpful to astronomers as the human eye perceives intensity logarithmically.

As the brightness scale has no upper or lower limit, bright celestial objects like Venus and Sirius can have negative apparent magnitudes. The zero-point for m was traditionally set to 0 for the star Vega (this is still true for visible to near-infrared wavelengths), currently different calibrations are used for different wavelength ranges, the reader could refer to Bessell (2005) for further details. Other magnitude systems have zero-points defined differently, like the AB magnitude system, which will be discussed further in Section 2.1.

While m and M_V are both measurements of luminosity, the combination of both measurements, with the help of theory or objects of known distances, can help us to gauge distances to more distant object. When the apparent magnitude of a star or galaxy is measured, M_V can be estimated if the distance d (in parsecs) to the object is known:

$$m - M_V = 5 \log_{10} \frac{d}{10}. \tag{17}$$

The term $\mu_d = m - M_V$ is also known as the *distance modulus*, at times a more convenient way to express distances to nearby galaxies and supernovae.

1.2.5 *The Hertzsprung-Russell Diagram*

Stars which are further away have parallaxes with high uncertainties, distances to these stars can be obtained and improved with the help of the properties of nearer stars that have been well studied and classified through the *Hertzsprung-Russell* (HR) diagram. The HR diagram was developed by Ejnar Hertzsprung and Henry Russell around 1910. It is a scatter plot of stars with their absolute magnitudes plotted against their spectral classification. In other words, this diagram shows the luminosity-temperature (or colour) relationship of stars. Here luminosity is defined to be the total amount of energy emitted by an object per unit time, an entity dependent on the object and not distance nor area.

Figure 2 shows an example of the HR diagram. The diagram exists in many forms, but the general rule is that stars with greater luminosity are near the top of the diagram, while stars with lower surface temperatures (redder stars) are nearer to the right. Perhaps a more popular form of it is the colour-magnitude diagram, which is the most common form seen in modern astronomy journal papers. The HR diagram inspired Arthur Eddington to develop ideas on stellar evolution (Eddington 1920), which led to the eventual discovery of nuclear fusion.

Evolutionary patterns relating mass, age and composition can be deduced from the HR diagram. For example, main sequence stars lie within a diagonal line sloping downwards in the diagram, they are hydrogen fusing stars. With their spectra measured, their absolute magnitudes can be estimated by apparent magnitude measurement, after correcting for interstellar extinction due to gas and dust (more on this in Section 2.1). This method of finding distances to main sequence stars with the help of the HR diagram is called *main-sequence fitting*. Stars in a star cluster are known to have roughly the same distances from Earth, thus when plotted their apparent magnitudes against temperature, we could compare it to a catalogue of main sequence stars plotted with their absolute magnitudes against temperature. When both plots are coincided, the vertical difference of points in the diagram would be the distance modulus, whereby the distance to the star cluster could be estimated. Recent work using main-sequence fitting include O'Malley et al. (2017) in determining absolute ages and distances to globular clusters with the help of data from Gaia. This method however has a lot of complications, the reader could refer to Pinsonneault et al. (2000) for further review on this topic.

With the help of the HR diagram, stars that are further than 10 kpc can also be measured via *spectroscopic parallax*[2], where the spectra of stars are measured and compared to properties of stars in the HR diagram. In fact the more commonly used method should be the *photometric parallax*, where the apparent magnitude of stars in multiple passband filters are used instead of their spectra, in which the Sloan Digital Sky Survey (SDSS) used this method to measure distances to 48 million stars as far as 20 kpc (Jurić et al. 2008). Photometric parallax is in fact just the 'photometric' version of spectroscopic parallax, just like how photometric redshift is to spectroscopic redshift (more on this in Chapter 3). The reader could refer to Sandage et al. (2016) on the development history of the photometric parallax method, which has been in use since the 1930s.

1.2.6 *Classical Cepheid Variables*

A *Cepheid variable* is a type of star that pulsates radially, its diameter and temperature varies periodically causing a change in brightness. They are known to have a well-defined and stable pulsation and amplitude. There are two main classes of Cepheids: *classical Cepheids* are more massive and younger population I stars, whereas *type II Cepheids* are older fainter Population II stars.

Classical Cepheid variables have masses between $4M_\odot$ to $20M_\odot$, up to 50 000 times more luminous than our Sun, and are generally bright giants of spectral class $F6 - K2$ (Turner 1996). Classical Cepheids have a characteristic asymmetric light curve, with a rapid rise to maximum

2 Despite having the word 'parallax' in its name, it is worth noting that this method does not make use of trigonometrical parallax at all. The same applies to the term 'photometric parallax'.

light followed by a slower fall to minimum due to the phase difference between radius and temperature variations.

The first classical Cepheid variable star, the δ Cephei was discovered by John Goodricke (Goodricke and Bayer 1786), but it was only until 1912 when Henrieta Leavitt found a period-luminosity relationship for classical Cepheids (Leavitt and Pickering 1912). Leavitt studied 25 Cepheid stars in the Magellanic clouds and found that the brighter the Cepheid, the longer its period. Their periods are found to be regular and have uniform functions of brightness, and since they are reasonably abundant and very bright, the period-luminosity relation of classical Cepheids allows them to be cosmic yardsticks to distances up to 50 Mpc. Classical Cepheids also allowed Hubble to conclude that M31 (Andromeda) was actually an external galaxy instead of a smaller nebula within the Milky Way, which was commonly thought to be at that time (Hubble 1929b).

The period-luminosity relation of classical Cepheid variable stars depend on the filters and data used, an example from a recent paper by (Fritz Benedict et al. 2007) suggested the following empirical equation:

$$5 \log_{10} d = V + 3.34 \log_{10} T - 2.45(V - I) + 7.52, \tag{18}$$

where T is the period, V and I are magnitudes of the different bandpasses used. Therefore with the apparent magnitudes and periods of Classical Cepheids known, the distances to these stars can be estimated.

However, various uncertainties and unresolved issues in Cepheids made it a constant and active debate in literature. Recent research on the reliability of this method include the study of the impact of metallicity (Scowcroft et al. 2009) and blending (Mochejska et al. 2000) on the period-luminosity relation. Despite that, Classical Cepheids still play a major role as a traditional *standard candle* in the cosmic ladder.

1.2.7 *Maser*

The *microwave amplification by stimulated emission of radiation* (maser) is the microwave counterpart of laser, it is a highly coherent and monochromatic electromagnetic wave in the microwave frequency. In the context of astrophysics, masers are not referred to the artificially engineered masers, they are instead referred to the naturally occurring source of stimulated microwave spectral line emission, which could be produced by hydroxyl radicals (OH), water (H_2O) and many other molecules.

Astrophysical masers are generally time-varying, polarised, bright and have very narrow line widths (Claussen 2004). They are found in star forming regions, in circumstellar shells of dying stars, supernova remnants and nuclei of active galaxies. Astrophysical masers are formed when amplification of microwave radiation takes place in a region of gas rotating around a massive source, high energy radiation from the source pumps the molecular rotational and vibrational transitions of the gas molecules, population inversion then occurs and masers are emitted, with the rotation around the source causing Doppler broadening of the spectral line. Maser emission can be distinguished from thermal emission by their brightness temperature, reaching $> 10^9$ K, which is at least 6 orders higher than thermal emissions (Humphreys 2011).

Astrophysical masers were first observed in the 1960s by Weaver et al. (1965) to be of unknown origin, but later found to be emission lines due to OH molecules. Since then masers originating from different molecules were found, and it also was found that these masers were produced when X-ray emission from hot gas near black holes stimulate these molecules. Masers have been used to study turbulence in young stars and the structures of galaxy disks (Strelnitski 1997).

Masers were first used to determine distances to nearby galaxies in the late 1990s. Herrnstein et al. (1999) used the Very Long Baseline Array of The National Radio Astronomy Observatory (NRAO) to obtain the distance to NGC4258, which also acted as the first direct evidence for a supermassive black hole in a galactic nucleus. This is possible due to the sharp and strong maser emission lines, angular distribution of maser features around accretion disks of supermassive black holes can be mapped with Very Long Baseline Interferometry (VLBI), and distances to these black holes can be estimated geometrically as masers orbit around them over a period of years. Masers allow distances to nearby galaxies be measured up to 3% accuracy (Herrnstein et al. 1999).

The maser distance estimation method is entirely geometric and independent of luminosity calibrations, thus it is of particular interest in current astrophysics research. Current active research on astrophysical masers are conducted by the Megamaser Cosmology Project (MCP, Reid et al. 2009). With more accurate distance measurements to nearby galaxies, MCP hopes to lower the uncertainty of H_0 to about 1%.

1.2.8 *The Cosmological Redshift*

Until now we have mostly dealt with distances to planetary objects, stars and nearby galaxies. From this point onward, we have arrived at the part of the cosmic distance ladder in which we have to deal with intergalactic distances involving galaxies, supernovae and quasars. In fact, this will be the basis of distance measurements relevant to this research.

In order to measure distances to distant galaxies and supernovae, we need to define the *cosmological redshift*, which is the elongation of the wavelength of electromagnetic radiation measured for an object moving further away from us. As we have established earlier in Section 1.1, the consequences of living in an expanding universe is that at very large distances, galaxies are moving at great speeds away from us which result in a significant shift in wavelength of light we receive. This is known as the *cosmological redshift* z, defined mathematically to be

$$z = \frac{\lambda_{\text{obs}} - \lambda_{\text{em}}}{\lambda_{\text{em}}}, \tag{19}$$

where λ_{obs} is the wavelength of light received by the observer and λ_{em} is the wavelength of light emitted by the source.

Cosmological redshift is not to be confused with the Doppler redshift of light in special relativity (when sources and observers are moving relative to one another), or the gravitational redshift in general relativity (observers at different locations in a gravitational field measure different wavelengths), since it happens due to the expansion of space. Early discussions on astronomical redshifts were understood as Doppler redshifts mainly until the Friedman equations were established (Friedman 1922), and when Hubble showed the linear correlation between

galaxy redshift and distance (Hubble 1929a). In fact the Doppler redshift is indeed observed in astronomy mainly through the Doppler shift associated with the random peculiar velocities and random motions of galaxies in clusters (the Fingers of God effect, Jackson 1972) or through the coherent motions of galaxies falling inwards towards an assembling cluster (the Kaiser effect, Kaiser 1987), which result in what we know as redshift-space distortions.

Gravitational redshifts, on the other hand, are observed as well and happens through gravitational lensing and redshift of photons from the CMB (Sachs-Wolfe effect, Sachs and Wolfe 1967). Redshift-space distortion is a very important cosmological probe in its own right, and it introduce redshift uncertainties especially significant for very low redshift objects, but for the purpose of this work we shall not discuss this further. From this point onward any mention of 'redshift' shall refer to the cosmological redshift unless otherwise stated.

To construct a relationship between z and the scale factor $a(t)$, we use the geodesic equation for a light wave, $ds^2 = 0$. From the FLRW metric (Equation 2), we find

$$\frac{\lambda_0}{\lambda(t)} = \frac{a_0}{a(t)}, \tag{20}$$

where λ_0 and a_0 are the wavelength observed and scale factor at present time (see Peacock 1999, for the complete derivation). Defining $\lambda_{\rm obs}$ to be the wavelength of light from a galaxy observed in the present and $\lambda_{\rm em}$ be light emitted from that galaxy at a certain time in the past, we compare this this with Equation 19 to get

$$a(z) = \frac{1}{1+z}, \tag{21}$$

where we have previously defined $a_0 = 1$. With this relationship we can now rewrite the Hubble parameter (Equation 12) in terms of redshift:

$$H(z) = H_0\sqrt{\Omega_{\rm m,0}(1+z)^3 + \Omega_{\Lambda,0}}. \tag{22}$$

With this equation we can now define the *radial comoving distance*, which is the distance between two nearby objects in the universe which would remain constant with epoch if the two objects are moving with the Hubble flow (or, the proper distance divided by the scale factor a). The comoving distance $d_{\rm C}$ along the line-of-sight can be written as follows:

$$d_{\rm C}(z) = c\int_0^z \frac{1}{H(z')}\,dz' = \frac{c}{H_0}\int_0^z \frac{1}{\sqrt{\Omega_{\rm m,0}(1+z')^3 + \Omega_{\Lambda,0}}}\,dz'. \tag{23}$$

Here we integrate from $z = 0$ (the redshift at present) where the light is observed to the redshift z when it was emitted. The transverse comoving distance (two events at a certain redshift z but separated on the sky by some angle $\delta\theta$) can be written as $d_{\rm C}(z)\delta\theta$ for a flat universe. The current physical distance to that object can be measured by multiplying the scale factor of today to the comoving distance, however we are more interested in the redshift and comoving distance than the physical distance itself.

The comoving distance is not directly measured, but can be measured through the luminosity L of the object. When the absolute luminosity of an object is known we can measure the flux F

and determine its *luminosity distance*. Note here that flux is the total light energy intercepted by the detector divided by the area of the detector, which is a value dependent on area and distance to the object. The luminosity distance $d_{\mathrm{L}}^2(z)$ can be defined as follows:

$$d_{\mathrm{L}}^2(z) = \frac{L}{4\pi F}. \tag{24}$$

Relating luminosity distance with comoving distance, it can be shown that

$$d_{\mathrm{L}}(z) = (1+z)d_{\mathrm{C}}(z) = \frac{c(1+z)}{H_0}\int_0^z \frac{1}{\sqrt{\Omega_{\mathrm{m},0}(1+z')^3 + \Omega_{\Lambda,0}}}\,dz', \tag{25}$$

since $F \propto \frac{1}{d_{\mathrm{C}}^2}$ and $F = F_0(1+z)^2$ (see Copeland et al. 2006, for full derivation). $d_{\mathrm{L}}(z)$ can also be defined by the absolute and apparent magnitudes, which in fact is a rearrangement of Equation 17:

$$d_{\mathrm{L}} = 10^{\frac{(m-M_V+5)}{5}}. \tag{26}$$

Finally, we could also define the *angular diameter distance* d_{A}, which is the ratio of an object's physical transverse size to its angular size. In a flat universe, we can relate d_{A} to d_{C} through the equation

$$d_{\mathrm{A}} = \frac{a(z)\,d_{\mathrm{C}}(z)\,\delta\theta}{\delta\theta} = \frac{d_{\mathrm{C}}(z)}{1+z}. \tag{27}$$

The angular diameter distance is known not to increase indefinitely. Assuming an unchanging galaxy with a fixed size, we expect the ratio between the actual size of a galaxy to its angular size to increase indefinitely as the galaxy gets further away from us. However, in an expanding universe this is more complicated, since physical sizes decrease with time and distance, the unchanging galaxy with a fixed size will have its angular size increase up until about $z = 1.5$, and it would decrease beyond that, as visualised in Figure 3. In simple words, objects at higher redshifts have larger perceived sizes than they should. The reader could refer to Hogg (1999) for a comprehensive discussion on the different distances mentioned in this section.

1.2.9 *Faber-Jackson Relation and the Fundamental Plane*

Having cosmological redshifts defined, we can finally begin a discussion on distance measurements to galaxies. The Hubble sequence classified three main kinds of galaxies based on their shapes: ellipticals, spirals and lenticular galaxies. Elliptical galaxies are well understood, they have surface brightness profiles which can be completely described by their radii and intensity (more on this in Section 2.1.6). The *Faber-Jackson relation* (Faber and Jackson 1976) is an empirical relation between the luminosity L and *velocity dispersion* σ_0 of elliptical galaxies, mathematically

$$L \propto \sigma_0^n, \tag{28}$$

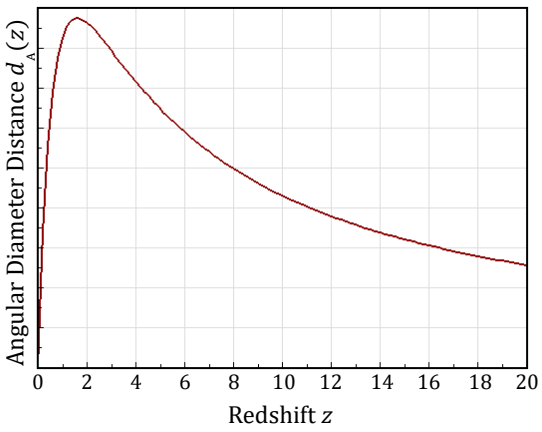

Figure 3: Plot of angular diameter distance (d_A) v.s. redshift (z). Figure created at http://www.icosmos.co.uk.

where n is an index observed to be about 4, assuming a constant mass-luminosity ratio. The velocity dispersion of a galaxy is the statistical dispersion of velocities about its mean, and this can be observed via the broadening of emission lines.

Faber argued that the motivation for this relation was due to gravitational collapse (Peacock 1999). In fact, we can almost derive this relationship. Consider the gravitational potential of a mass distribution, using the virial theorem and a constant light to mass ratio, we can find

$$L\mu \propto \sigma_0^4, \tag{29}$$

where μ is the surface brightness of the elliptical galaxy, and we see that an assumption of constant μ gives rise to the Faber-Jackson relation. This means that once the velocity dispersion of an elliptical galaxy is known by means of spectroscopy, its absolute magnitude can be derived. Together with the apparent magnitude the distance to the galaxy can be estimated.

However, in most cases μ is not constant, and since L is proportional to the effective radius of the galaxy ($L = \mu_0 \pi R_0^2$), therefore an empirical relationship based on Equation 29 can be formulated:

$$R_0 \propto \mu^m \sigma_0^n, \tag{30}$$

which is known as the *fundamental plane* for elliptical galaxies. The fundamental plane was first mentioned by Terlevich et al. (1981) and later discussed in a more detailed manner by Djorgovski and Davis (1987). This relation tells us that elliptical galaxies are self-gravitating systems, and they have almost constant mass-to-light ratios (Peacock 1999).

Knowing that the many characteristics of a galaxy are correlated, the usefulness of these correlations motivated such an empirical equation to allow the distance of the galaxy to be calculated. In this case, when the emission line width and apparent magnitude of the elliptical galaxy is known, the effective radius can be estimated, and with approximations of angular size the distance to the galaxy can be found. A similar relation is the D_n-σ_0 relation put forward by Dressler et al. (1987), where D_n is the diameter within which the mean surface brightness is 20.75.

A recent work on the fundamental plane was put forward by Saulder et al. (2013) where they calibrated the fundamental plane with data from the SDSS MGS. They carefully selected a group of elliptical galaxies with the help of the morphological classification of Galaxy Zoo (Lintott et al. 2011), and fit the data into the following equation,

$$\log_{10}(R_0) = \alpha \log_{10}(\sigma_0) + \beta \log_{10}(I_0) + \gamma \tag{31}$$

where α, β and γ are free parameters to be fit. R_0 is the effective physical radius of the galaxies, while $\mu_0 = \log_{10}(I_0)$ is the surface brightness, which can be written as

$$\mu_0 = m - A - K + Ez_c + 2.5\log_{10}(2\pi s^2) - 10\log_{10}(1 + z_c), \tag{32}$$

where m is the apparent magnitude, A the extinction correction, K the K-correction, E an evolutionary parameter, z_c the CMB rest frame redshift, s the circularised angular radius, while the last term corrects for cosmological dimming of surface brightness (more on A, K and cosmological dimming in Section 2.1.5). After fitting the data, they found an accuracy of 15% for the fundamental plane as a distance indicator. We will revisit this equation in Section 6.2.3.

It is worth the mention that there is in fact a spiral galaxy version of the Faber-Jackson relation, known as the *Tully-Fisher Relation* (Tully and Fisher 1977). This is an empirical relationship for spiral galaxies, establishing a relationship between its luminosity and rotational velocity. The Tully-Fisher relation is known to be not as tight as the Faber-Jackson relation, and McGaugh et al. (2000) eventually showed that the Tully-Fisher relation could be tighter when luminosity is replaced by the galaxy's total baryonic mass, resulting in the *Baryonic Tully-Fisher Relation*.

1.2.10 *Type Ia Supernovae*

Supernovae are cataclysmic nuclear explosions in stars, they are very luminous objects which are able to outshine an entire galaxy, but later fade off slowly with time. A *Type Ia Supernova* (SNIa) is a subclass of supernovae, it does not contain hydrogen, and presents a singly ionised silicon (Si II) line at 615 nm near peak light. SNIa are formed in binary star-white dwarf systems, when the gas from a star is transferred to the white dwarf due to gravity, the white dwarf accretes too much material and explodes as it reaches the *Chandrashekar limit* of $1.4M_\odot$, which is the maximum stable mass of a white dwarf star.

Since all SNIa explode at almost the same conditions, it is supposed that the differences in peak luminosities of SNIa are correlated with how quickly their light curves decline after maximum light. The peak of the light curve of all SNIa reaches a consistent luminosity with absolute magnitude $M_V = -19.3$, therefore they are said to be *standard candles*: they have a known standard luminosity which aids us to calibrate the cosmic distance ladder. SNIa is a strong constraint of dark energy, it is used for the determination of distances to distant galaxies.

Since we know the absolute luminosity L of SNIa, we can measure its apparent magnitude to calculate its distance. We substitute $M_V = -19.3$ and its apparent magnitude m in Equation 17,

and using Equation 25, we can finally express the distance modulus of supernovae in terms of its redshift:

$$\mu_d(z) = 5\log\left[\frac{d_L(z)}{10\,\mathrm{pc}}\right] = 5\log\left[\frac{c(1+z)}{H_0}\int_0^z \frac{1}{\sqrt{\Omega_{m,0}(1+z')^3+\Omega_{\Lambda,0}}}\,dz'\right]. \tag{33}$$

In other words, once we obtain the apparent magnitude of the SNIa by fitting its light curve, we can get a distance-redshift relation numerically assuming ΛCDM cosmology, and hence determine the redshift of its host galaxy.

1.2.11 *Baryon Acoustic Oscillation*

Another way of measuring extra-galactic distances is that if the actual size of an object is known, we can find the distance to the object based on its angular size θ. An object with a known standard physical size is called a *standard ruler*.

Baryon Acoustic Oscillation (BAO) is the regular periodic fluctuation of matter density in visible baryonic matter, caused by acoustic waves which existed in the early universe. According to the Big Bang model, photons were not decoupled from baryons and electrons prior to the Recombination era ($z > 1100$), the photons in the plasma of baryons and electrons interacted through Thomson scattering, an elastic scattering process. This created a pressure on that medium through which acoustic waves travelled at the speed of sound c_s,

$$c_s = \sqrt{\frac{\partial P}{\partial \rho}} = \frac{c}{\sqrt{3\left(1+\frac{3\rho_b}{4\rho_\gamma}\right)}}, \tag{34}$$

where P is pressure, ρ_b and ρ_γ are the densities of the baryons and photons respectively. Since dark matter is not affected by electromagnetic waves, it becomes the source of the overdensity, while the pressure asserted by the photons and the gravitational pull by the dark matter on the baryons caused an oscillation of densities. During the recombination era, photons are decoupled from the baryons and free-stream throughout the universe, which becomes the CMB we observe today. In fact, the gravity-driven acoustic oscillations of the coupled photon-baryon fluid is seen through the peaks and troughs in the angular power spectrum of the CMB temperature anisotropy. The vanishing pressure on the medium made the travelling baryons stall at a shell of a certain comoving distance r_{drag} from the initial overdensity where they were created:

$$r_{drag} = d_C(z_{drag}) = c_s\int_{z_{drag}}^{\infty}\frac{1}{H(z')}\,dz' \approx 150\,\mathrm{Mpc}. \tag{35}$$

This is also known as the *sound horizon* (Eisenstein et al. 2005). The integral goes to $z = \infty$ which refers the beginning of the universe. In simple words, this caused an excess of galaxy clustering which are separated by the distance of 150 Mpc, more than other distances caused by

the baryon oscillation. Since we could measure the angle θ subtended by the galaxies, with an overdensity of a fixed distance, we can find its redshift by the following formula:

$$\theta = \frac{r_{\text{drag}}}{d_{\text{A}}(z)}. \tag{36}$$

In fact we could also measure its radial size δz, the difference in redshift between the nearest and the furthest part of the standard ruler (assuming a spherical object):

$$\delta z = \frac{r_{\text{drag}} H(z)}{c}, \tag{37}$$

in which we could also measure the Hubble parameter. However it is not possible to estimate single measurements like this directly, it could instead be done by studying separations between a large set of galaxies: in terms of small 'wiggles' on the matter power spectrum or small 'bumps' on galaxy auto-correlation functions. BAO also acts as a constraint to dark energy, since the spatial distribution of galaxies at different epochs differ between dark energy models.

Eisenstein et al. (2005) detected this preferred distance by measuring the 3D redshift-space correlation function $\xi(s)$ with the SDSS LRG spectroscopic sample. The Baryon Oscillation Spectroscopic Survey (BOSS), which is one of the major surveys in SDSS-III has determined this angular diameter distance at redshifts $z = 0.32$ and $z = 0.57$ with a precision of 1% (Anderson et al. 2014). In the recent years the WiggleZ survey applied the reconstruction of the baryonic acoustic feature technique and showed significant improvements in distance measurements (Kazin et al. 2014). The most recent results of Planck (Ade et al. 2016) has calculated best-fit values of the sound horizon $r_{\text{drag}} = 147.6 \pm 0.4$ Mpc and $z_{\text{drag}} = 1059.6 \pm 0.5$ respectively, determined by the structure of the acoustic peaks in the CMB. The BAO remains a powerful low-redshift probe, as it is limited mostly by statistical uncertainties rather than systematics.

1.2.12 *Other Methods*

Galaxies and quasars which are further and fainter than what has been discussed would face more difficulty obtaining redshifts. This leads us to the next 2 chapters, where we would discuss the use of spectroscopic and photometric redshift estimations to gauge their distances from Earth. We end this chapter by mentioning 4 other distance measurement methods, in which their discussions were not long enough to earn themselves a section each, yet interesting enough for the reader:

- *Eclipsing Binaries* – Eclipsing binary stars are binary systems where their orbit plane lies close in the line of sight, which allow them to undergo periodic mutual eclipsing. We could use them to directly measure distances to external galaxies, like the Large Magellanic Cloud (LMC, Guinan et al. 1998). The reader could refer to Bonanos (2006) for more details.
- *Dynamical Parallax* – For stars which are slightly further, instead of using the parallax of Sun, this method uses the parallax between the orbits of two rotating binary stars, the apparent magnitude and the period of the rotation to obtain its distance (Russell 1928). This is done with the help of the mass-luminosity function ($L \propto M^a$, Kuiper 1938) for binary stars, where a is a value between 1 to 6, depending on the mass of the star. Recent

use of dynamical parallax include the work of Caballero (2008) to obtain the distance to σ Ori AB.

- *X-Ray or Gamma Ray Bursts* – The progenitors of these very energetic radiation sources are still being studied, but the main consensus is that X-ray bursts arise from X-ray binary stars (neutron star) with periodic and rapid increases in luminosity that peak in the X-ray wavelength, while gamma ray bursts may have come from hypernovas. These objects are very far away, and some have suggested the possibility of them being standard candles for redshifts $z > 3$. Thus far results have been inconclusive: Kuulkers et al. (2003) thinks that there is a slight positive chance that X-rays may be standard candles, Gendre et al. (2008) is very confident that it could, while Lin et al. (2015) suggested that it is insensitive to cosmological parameters. We expect to see more conclusive results with better X-ray and gamma ray detectors being built in the future.

- *Gravitational Waves* – Gravitational waves from spiralling compact binary neutron stars or black holes have their amplitudes and shapes strongly depend on the mass of the system. These can be computed when a waveform is observed, which makes gravitational waves a *standard siren* of known loudness. Early studies have been conducted to cross-validate with optical observations (LIGO Scientific Collaboration and The Virgo Collaboration et al. 2017). We are probably just at the dawn of a new distance measurement method.

2

GALAXY PHOTOMETRY, MORPHOLOGY AND SPECTROSCOPY

When I consider Your heavens, the works of Your fingers, the moon and the stars, which You have set in place, what is man that You are mindful of him, the son of man that You care for him?

Psalms 8:3-4

2.1 PHOTOMETRY

In the previous chapter we have discussed how redshifts are important distance indicators for celestial objects which are very far away, and we have seen that they could be determined using standard candles, standard rulers and empirical relations. However, in order to continue our discussion on another method for redshift determination (i.e. spectroscopic redshift), we need to first understand *astronomical photometry* – the technique of measuring photon flux of astronomical objects over certain wavelength intervals – which is a main ingredient of measuring spectroscopic and photometric redshifts. We have briefly defined the absolute and apparent magnitude of stars in Section 1.2.10, and we will discuss further in this chapter on fluxes and magnitude systems.

2.1.1 *Galaxy Luminosity Function*

Before we dive into the technicalities of measuring flux and magnitudes, let us first discuss the distribution of celestial objects via their luminosity. The *galaxy luminosity function* gives the number of galaxies up to a certain luminosity (the light energy released by the galaxy per unit time) in a certain volume. It is found that the galaxy luminosity distribution follows a power law at low luminosities, and truncates exponentially at high luminosities. Schechter (1976) parametrized the galaxy luminosity function, thus the number of galaxies greater than a certain luminosity $L_{\rm lim}$ can be represented by the equation

$$N(L_{\rm lim}) = \int_{L_{\rm lim}}^{\infty} \phi_\star x^\alpha e^{-x}\, dx, \tag{38}$$

where α is a power law index for low luminosity; $\phi_\star$ is a normalisation factor representing the number of galaxies per unit volume, while $x = \frac{L}{L_\star}$, $L_\star$ is the characteristic galaxy luminosity where the power-law form of the function cuts off. This empirical equation is such that parameters α, $\phi_\star$ and $L_\star$ can be determined by data and usually measured according to specific

bandpasses, and in fact the values are calculated to be different for elliptical, spiral and irregular galaxies (Dahlen et al. 2005). x can also be written in terms of absolute magnitudes where $\frac{L}{L_\star} = 10^{-0.4(M-M_\star)}$. Given an absolute magnitude as input, the galaxy luminosity function essentially returns the abundance of objects brighter than that absolute magnitude. Although this form of luminosity function is widely used by many, it is however worth pointing out that if $L_{\rm min} \to 0$, $N(L_{\rm lim})$ would diverge. The true luminosity function must drop at low L, but this has not been observed yet.

The luminosity function is used to study the properties of stars in clusters or the galaxies in the Local Group, it provides information about the power spectrum of the primordial density fluctuations and star formation (Jerjen 2001). The galaxy luminosity function can also used to generate simulated galaxy samples, as we will see in Section 4.2.2.

2.1.2 *Flux and Magnitude Systems*

In Section 1.2.8 we have briefly mentioned the relationship between magnitude and flux. *Flux* is the amount of electromagnetic radiation energy we receive from the object per unit area and time. In astronomical photometry the flux of an object is usually measured through a *filter*. The filter, as its name suggests, would block out light which is outside of the selected bandpass wavelength. Generally filters that are larger than a width of 300 Å are known as *broadband filters*, and narrowband filters if otherwise. A *photometric system* is a set of filters with well defined bandwidths, we will see more on this in Section 2.1.3.

Consider some light passing through a filter. The probability that a photon with wavelength λ' passes through a bandpass with median wavelength λ is the filter response, $R_\lambda(\lambda')$. Therefore the total flux F passing through a bandpass is

$$F = \int_0^\infty f_\lambda(\lambda') R_\lambda(\lambda')\, d\lambda', \tag{39}$$

where $f_\lambda(\lambda')$ is the spectral energy distribution (SED) of the galaxy (flux per unit wavelength). The total filter response $R_\lambda(\lambda')$ is a product of a few terms,

$$R_\lambda = T_{\rm CCD} T_{\rm opt} T_{\rm atm} \tag{40}$$

where the T's representing the transmission curves due to the quantum efficiency of the charge-coupled devices (CCDs), telescope optics (opt) and observatory atmospheric conditions (atm). $T_{\rm atm}$ is only relevant when observed from the ground, when observed from space we would have $T_{\rm atm} = 1$.

Similar to Equation 16, we can define the apparent magnitude by the following equation,

$$m = -2.5 \log \frac{F}{F_0} = -2.5 \log F + m_0, \tag{41}$$

where F_0 and m_0 define the zero-point flux and magnitude respectively. The zero-point magnitude depends on the magnitude system used. For example in the *Vega magnitude system*, the magnitude zero-point for each band is set such that the star Vega would have zero magnitude in

all bands, resulting in Vega having zero colours (Bessell 2005). This results in different zero-point values for different bands used.

A more common magnitude system used today is the *AB magnitude system* (with AB referring to 'absolute'). Rather than using a stellar spectrum as a reference, it is based on a hypothetical reference spectrum which has constant flux per unit frequency interval. The reference flux in the AB system is defined to be $F_0 = 3631$ Jy (Oke and Gunn 1983), therefore $m_0 = 8.9$ if the flux density is measured in Jansky (Jy); $m_0 = 48.6$ if measured in erg s^{-1} cm^{-2} Hz^{-1}; and $m_0 = 23.9$ if measured in nanomaggies or μJy, as per SDSS and the Cosmic Evolution Survey (COSMOS). A zero magnitude in the AB system corresponds to the flux density of Vega at the effective wavelength of the Johnson V band (at about 5500 Å). The AB magnitude system was originally a monochromatic system (i.e. flux of a single wavelength), but as shown in Equation 41, this can be used to get apparent magnitudes of broadband fluxes too.

SDSS however does not use the AB system, instead it uses the *asinh magnitude system* (hyperbolic sine), which follows the following equation (Lupton et al. 1999):

$$m = -\frac{2.5}{\ln 10}\left[\sinh^{-1}\left(\frac{F}{2bF_0}\right) + \ln b\right] \tag{42}$$

where F_0 is the zero-point flux as defined in the AB system, while b is a softening parameter, which is the 1σ noise of the sky in a point spread function (PSF) aperture in $1''$ seeing. asinh magnitudes are identical to AB magnitudes at high signal-to-noise ratio (denoted as $^S/_N$ throughout this paper), in contrast it behaves reasonably well at low $^S/_N$ and even at negative fluxes (see Section 2.1.4), when the AB magnitude does not.

2.1.3 *Photometric Systems and Colours*

In the previous section we have mentioned that in astronomy fluxes are measured through filters, and a well defined set of filters with different wavelengths are known as *photometric systems*. Photometric systems are defined based on the survey of interest, e.g. an optical survey would have filters in the optical range. The oldest photometric system is the *UBV photometric system* (Johnson and Morgan 1953), it is also known as the Johnson-Morgan system. This is a set of three filters with full width at half maximum (FWHM) between 660 Å and 940 Å: ultraviolet U with effective wavelength midpoint 3500 Å, blue B with 4350 Å, and visual V at 5500 Å. The filters are at the blue end of the spectrum due to the bias that photographic films have for those colours, while the short wavelength cutoff of the U filter is defined mainly by the terrestrial atmosphere. The UBV system was later extended to the redder spectrum to become the Johnson-Cousins $UBVRI$ system (Cousins 1980; Bessell 2005).

Other photometric systems exist, the most notable being the SDSS *ugriz photometric system* (Fukugita et al. 1996) in which 158 standard stars form the basis for its photometric calibration (Smith et al. 2002). The SDSS *ugriz* system uses the same zero-point as Oke and Gunn (1983) and is made to have wide bandwidths to ensure a high efficiency for faint object detection and also to cover the entire accessible optical wavelength range. The division between the *gri* passbands were designed such that the strongest night-sky lines, namely the O I 5577 Å line, Hg I 4358 Å line and Hg I 5460 Å line fall in the gaps of those bands (Thuan and Gunn 1976) and thus

not contaminate the photometry. Many other systems were modelled after the Johnson-Cousins *UBVRI* or the *ugriz* system, thus we do not discuss these further, but the reader could refer to Bessell (2005) for a good summary on many current widely used photometric systems.

Within a photometric system we can define the *colour index* of a galaxy. It quantifies how blue or red an object is. The colour of an object is defined to be the difference between two apparent magnitudes m_a and m_b where $\lambda_a < \lambda_b$. The more negative the value, the bluer the object. Galaxy colours tells us about their temperature and age: red galaxies tend to be ellipticals, while blue galaxies tend to be spirals or star-burst galaxies with high metallicity. With more than two colours we can plot *colour-colour diagrams*, which are very useful for detecting colour outliers (e.g. identifying quasars from stars).

2.1.4 *Detectors*

The design of a detector is important to measure accurate fluxes from celestial objects. In the past images have been recorded using *photographic plates*, which are light-sensitive emulsion of silver salts coated on thin glass plates. Later it was replaced by photographic paper which were easier to handle, but modern telescopes record measurements using CCDs.

When light is directed onto the focal plane of the telescope covered with CCDs, the number of incident photons N_{obj} on the CCD is counted, by converting them into electric charges through the photoelectric effect, and the accumulated charge will be measured and digitalized by applying weak electric fields. The flux of the object is measured using the fact that it is proportional to $\frac{N_{\text{obj}}}{t_{\text{exp}}}$ where t_{exp} is the exposure time of the telescope (Craig et al. 2014). The apparent size of the object is translated by the pixel scale from the number of pixels the image occupies. The flux measured will be calibrated with standard stars corrected for atmospheric absorption of light, and the choice of the standard stars and calibration method varies from survey to survey.

During the measurement process, there are a few sources of noise that need to be removed from the measured flux values. *Thermal noise* from the detector may add to the photon counts in the CCD array when the time exposure is high, therefore on top of keeping the detector cool and dry, several exposures will be taken with the CCD shutter closed to produce a *dark frame*, and this will be subtracted from the image to remove sensor noise or to account for defective pixels (this is also known as *flat-field correction*). On the other hand, *sky noise*, which is the incoming light from an apparently empty part of the night sky adds to the flux as well, a *background frame* which is taken by exposures on empty parts of the sky has to be subtracted from the image too. As mentioned in Section 2.1.2, some very faint objects with extremely low flux values may end up having negative fluxes when sky and background noises are subtracted from it, which when translated into magnitudes become a *nondetection*.

As the celestial objects are constantly moving across the sky, blurred or elongated images may be recorded in long exposures despite having the telescope on tracking. An *autoguider* can be mounted on a finderscope to take short exposures of an area of sky near the object, and the apparent motion can be corrected. Another method to solve the problem is by *drift-scanning*, where a CCD is used to make a fixed telescope behave like a tracking telescope and follow the motion of the sky. The charges in the CCD are read in a direction parallel to the motion of the

sky at the same speed, this technique is used by SDSS (Gunn et al. 2006) to image a larger region of the sky than its normal field of view.

Finally, ground-based telescopes are also affected by *seeing*, which is the blurring of celestial objects due to the variation of optical refractive index caused by turbulent mixing in the Earth's atmosphere. Seeing is the best possible angular resolution an optical telescope can achieve, it is characterised by the FWHM of the optical intensity across the PSF of the image seen through the atmosphere, usually measured in arcseconds. For example, SDSS has a median seeing of 1.4″ in the *r*-band (Hogg et al. 2001) while CS82 has a mean seeing of 0.6″ in the *i*-band (Moraes et al. 2014). Seeing is a highly variable quantity and is different according to place and time, good seeing can be achieved if observations are done on clear nights and at high altitude. Seeing can also be improved using *adaptive optics*, for instance, by using a wavefront sensor and deformable mirrors or using artificial guide stars. The reader could refer to Chromey (2010) for more information on how these methods work.

The 'quality' of a flux measured can be numerically presented by the *signal-to-noise* ($^{S}/_{N}$) of the flux. For ground-based telescopes, the magnitude can be measured by the number of photon counts $N_{\rm obj}$ at a certain exposure time, while the noise is given by adding quadratically three components, namely: the Poisson noise of the signal coming from the object ($\sqrt{N_{\rm obj}}$), the Poisson noise of the signal coming from the local intrinsic sky brightness ($\sqrt{N_{\rm sky}}$), and the Gaussian noise introduced by the electronics in the read-out system $\sigma_{\mathcal{N}}$. Given the number of exposures $n_{\rm e}$, $^{S}/_{N}$ can be written as the following equation (Martí et al. 2014):

$$^{S}/_{N} = \frac{N_{\rm obj}}{\sqrt{N_{\rm obj} + N_{\rm sky} + n_{\rm e}\sigma_{\mathcal{N}}^2}}. \tag{43}$$

For bright galaxies, $N_{\rm obj}$ would dominate the equation, thus $^{S}/_{N} \propto \sqrt{N_{\rm obj}}$; however for faint galaxies, $N_{\rm sky}$ would dominate and therefore $^{S}/_{N} \propto N_{\rm obj}$.

$^{S}/_{N}$ can also be represented by the ratio of the total flux F measured to the noise introduced $\sigma_{\rm F}$, therefore $^{S}/_{N} = \frac{F}{\sigma_F}$. This can be related to the *magnitude error* σ_m,

$$\sigma_m = |(m + \sigma_m) - m| = \left|-2.5\log_{10}(F + \sigma_F) + 2.5\log_{10}(F)\right| = 2.5\log_{10}\left(1 + \frac{\sigma_F}{F}\right). \tag{44}$$

For very small σ_F (or very high $^{S}/_{N}$), we can simplify Equation 44 by Taylor expansion, thus giving us

$$\sigma_m = 2.5\log_{10}\left(1 + \frac{1}{^{S}/_{N}}\right) \approx \frac{1.086}{^{S}/_{N}}, \tag{45}$$

in which for bright galaxies with small magnitude errors, $\frac{1}{\sigma_m}$ roughly translates to the $^{S}/_{N}$ of the object.

2.1.5 *Extinction and K-Correction*

Other than the various noise introduced by the detector or the sky, there are still three more important corrections for magnitudes measured in multi-band photometry. The first is *interstellar*

extinction, which is the absorption or scattering of light by mostly dust and gas of the interstellar medium between the celestial object observed and the observer. Since blue light is more strongly attenuated than red light, extinction causes *interstellar reddening*, where celestial objects would appear redder than expected. This is however different from redshift: extinction preferentially removes shorter wavelength photons from the spectrum without changing the position of spectroscopic lines.

Since extinction depends on the amount of dust in between the object and the observer (located mostly within our own Galaxy), its severity depends on the direction of the sky, e.g. the extinction at the galactic centre is so high that it could bring more than 30 magnitudes of extinction in the optical. Therefore to account for extinction we are essentially brightening the amplitude, we subtract the extinction term A_b from the apparent magnitude m_b, where b denotes the broadband filter used. Extinction is usually measured in terms of a *colour excess* $E_{B\text{-}V}$ in the B and V bands, it is the difference between the observed and intrinsic colour B-V, which explain the difference between the total extinction of an object in the B and V bands:

$$\begin{aligned} E_{B\text{-}V} &= (B\text{-}V)_{\text{obs}} - (B\text{-}V)_{\text{int}} \\ &= A_B - A_V. \end{aligned} \tag{46}$$

From Equation 46 we see that the colour excess can be determined even when the total extinction is not known in any of the photometric bands, it is a distance-independent measurement. The extinction map provided by Schlegel et al. (1998) provides values of $E_{B\text{-}V}$ for each line of sight of the sky; many modern surveys derive extinction correction values based on this. The extinction A_V for band V can calculated by multiplying the measured $E_{B\text{-}V}$ with the *extinction curve* R_V:

$$R_V = \frac{A_V}{E_{B\text{-}V}}. \tag{47}$$

The extinction curve depends on the galaxy and line of sight, but for earth-bound observations the mean value of the curve is commonly used and quoted, and for the Milky Way this gives $R_V = 3.1$. To measure the extinction in a different band b, one simply multiplies $E_{B\text{-}V}$ with a different R_b value (known as extinction factor F_f in Laigle et al. (2016), and a recent list of these values calculated for many well known bandpasses is found in Schlafly and Finkbeiner (2011).

The second important correction to magnitudes is the *K-correction*. *K*-correction corrects for the fact that the passband $R(\lambda)$ used to observe an object does not see the same spectrum $f(\lambda)$ of a source which has been redshifted by $1+z$, which was observed in a different band pass at its own rest frame. The *K*-correction can be calculated using the formula below (Hogg et al. 2002):

$$K = -2.5 \log \left[\frac{\int_0^\infty f_\lambda \left(\frac{\lambda}{1+z} \right) R(\lambda) \frac{\lambda}{1+z} \, d\lambda}{\int_0^\infty f_\lambda(\lambda) R(\lambda) \lambda \, d\lambda} \right]. \tag{48}$$

K-correction is only needed when measured in a bandpass of finite bandwidth, it is not needed if one uses the total bolometric magnitude or if one measures the light of an emission line. It is worth noting that *K*-correction can only be calculated when the redshift is known. If multi-band photometry exists, *K*-correction can be computed by fitting it against an theoretical or empirical

SED template (Blanton and Roweis 2007), approximated by a two-dimensional polynomial as function of redshift and colour for galaxies $z < 0.5$ (Chilingarian et al. 2010), or calculated using photometric redshifts (O'Mill et al. 2011). *K*-correction is needed in order to compare the photometric properties of galaxies at different redshifts in the correct frame.

Sometimes the discussion of *K*-correction would introduce an *E*-correction (Cole et al. 2001) which accounts for the passive evolution of galaxies. *E*-correction is motivated by the fact that *K*- and extinction-corrected apparent magnitude of galaxies changes in time due to stellar evolution. The simplest way to model *E*-correction is to make it proportional to redshift, yielding a term Ez to be added to the apparent magnitude m_b, and similar to K it is also a spectroscopic correction. Saulder et al. (2013) measured E to be 1.07 mag per z for the SDSS MGS sample.

2.1.6 *Magnitude Fits*

As the sensitivity of a detector varies with the type of detector and the wavelength of light it receives, it would be more useful if one specifies how the magnitude is measured. As mentioned earlier, integrated flux (or magnitude) of a celestial object can be computed directly by summing up the photon counts in a certain region, however the definition of the so-called 'certain region' is ambiguous. One easy way is to define the region by a fixed aperture, where it is the total flux of an object measured within its aperture (2″ or 3″ in diameter), however this is not ideal since the flux is dependent on object size, and light within this aperture may be contaminated by other sources. In this section we will define the different magnitude measures or fits and how they are measured. Since most of the photometry used in this paper come from SDSS, the definitions of the various magnitudes are specially defined in the context of SDSS unless stated otherwise.

First we start off with the *PSF magnitude*, the best magnitude to describe point sources. The PSF can be thought of as the extended 'blob' in an image which represents an unresolved object. In the ideal case where neither aberration nor turbulence is present, the telescope is said to be diffraction-limited, and its PSF has the form of an Airy disk. However, light from a point source like stars and quasars would appear on the focal plane with a certain size, due to aberration introduced by atmospheric turbulence and the diffraction of light by the telescope aperture. Therefore the image of a star can be seen as a convolution of the true object (a Dirac delta function) and the PSF. The PSF is modelled differently depending on the optics and sky conditions, though could be approximated by a Gaussian profile. A more accurate model is the Moffat distribution (Moffat 1969), where the intensity profile follows the following equation:

$$I(r, \alpha, \beta) = \frac{\beta - 1}{\pi \alpha^2} \left[1 + \left(\frac{r^2}{\alpha^2} \right) \right]^{-\beta}, \tag{49}$$

where α and β are seeing dependent parameters. The Moffat function is able to portray the wings of the PSF more accurately than the Gaussian profile. The PSF magnitude is then obtained by measuring the integrated flux F under it:

$$F = 2\pi \int_0^\infty I(r) r \, dr, \tag{50}$$

where $I(r)$ is the azimuthally averaged intensity profile of the object, depending on r the radius from the centre of the object[1], usually measured in arcseconds. A well modelled PSF is important to determine the Rayleigh criterion (angular resolution) of a telescope, this determines the closest distance between stars to be resolved by the telescope. Since atmospheric turbulence broadens the PSF, the diameter of a modelled PSF is a common way to quantify astronomical seeing.

Extended objects like galaxies are not well modelled by the PSF alone. The brightness of elliptical galaxies are better represented using the *de Vaucouleurs magnitude*, where the intensity profile $I(r)$ of the galaxy is modelled by the following equation (de Vaucouleurs 1948)

$$I(r) = I_0 e^{-7.67\left(\frac{r}{r_e}\right)^{\frac{1}{4}}}, \tag{51}$$

where I_0 is the intensity at the centre of the galaxy, and r_e the effective radius which is the radius of the isophote containing half the total flux. The brightness of disc galaxies, however are better represented by the *exponential magnitude*, which has an exponential intensity profile as follows (Freeman 1970):

$$I(r) = I_0 e^{-1.68\left(\frac{r}{r_e}\right)}. \tag{52}$$

The de Vaucouleurs and exponential intensity profile introduced by de Vaucouleurs and Freeman look similar and were observed to fit elliptical and disc galaxies well. In fact these two profiles are just special cases of the Sérsic (1963) profile:

$$I(r) = I_0 e^{-b_n\left(\frac{r}{r_e}\right)^{\frac{1}{n}}}. \tag{53}$$

b_n is a function that depends on n, it satisfies the equation $\gamma(2n; b_n) = \frac{1}{2}\Gamma(2n)$, where $\Gamma(2n) = \int_0^\infty x^{2n-1}e^{-x}\,dx$ is the Gamma function and $\gamma(2n, b_n) = \int_0^{b_n} x^{2n-1}\,e^{-x}\,dx$ is the lower incomplete Gamma function. If n is an integer, then $\Gamma(2n) = (2n-1)!$. b_n determined from the equation is in fact is what makes r_e the half light radius. n is the *Sérsic index* which controls the degree of curvature of the profile: the smaller the n, the less concentrated the profile and the steeper the slope at large radii. Thus, in some way, n can be used to determine the Hubble type of galaxies. As it can be easily shown, the de Vaucouleurs and exponential profile have $n = 4$ and $n = 1$ respectively, while a Gaussian profile has $n = \frac{1}{2}$.

Therefore the *Sérsic magnitude* is estimated by integrating the flux over a Sérsic profile. When fitting with the de Vaucouleurs, exponential or Sérsic profiles, other morphological parameters are output together with the magnitudes, these include the effective radius r_e, axis-ratio q (the ratio between the semi-minor and semi-major axes r_b and r_a of the galaxy when fitting an elliptical profile), the position angle, and the likelihood or χ^2 value of the model fit. It is also possible to fit more complicated models to a galaxy, e.g. using a de Vaucouleurs profile to model the bulge while an exponential profile to model the disc (or Sérsic fits for both), which are useful in characterising properties of lenticular galaxies. In this paper however we have not used magnitudes derived from these complex fits.

1 At this point of the thesis, the letter r can refer to either a radius or the r-band. Since the use of r the radius is so much more common than r the band in this thesis, and to avoid the use of uncommon symbols to further confuse the reader, the definition of r for the remainder of this chapter will mean radius, the word 'r-band' will always be present when referring to the band.

SDSS does not produce Sérsic magnitudes, however it has the *composite model magnitude* (or 'cmodel' magnitude), which was designed to produce a good estimate of the total flux of galaxies. This magnitude is based upon the best fitting positive linear combination of an exponential and de Vaucouleurs profiles, its composite flux can be represented by the following equation:

$$F_{\text{comp}} = f_{\text{deV}} F_{\text{deV}} + (1 - f_{\text{deV}}) F_{\text{exp}} \tag{54}$$

where f_{deV} is a coefficient with values between 0 and 1, showing the fraction of de Vaucouleurs flux (known as `fracDeV` in SDSS). The cmodel magnitude acts as a proxy for the best total flux indicator for galaxies and stars alike due to its good agreement with Petrosian magnitudes (explained in the next few paragraphs) of galaxies and PSF magnitudes of stars, however due to its dependence on a combination of model fits, it may not be a good colour indicator thus not suitable for colour sensitive procedures like photo-z estimation (Stoughton et al. 2002).

In order to get an optimal magnitude which best describes consistent colours across all bands, SDSS introduced the *model magnitude*, in which, slightly different from the *cmodel magnitude*, it only chooses the best of the de Vaucouleurs or exponential profile. The profile with the higher likelihood on the SDSS r-band is chosen. The same profile with the same effective radius r_{e} is then used for other bands (while allowing the amplitude to vary) after convolving with the appropriate PSF in each band. With this in state, the colour index between different bands for a particular galaxy is measured through a consistent aperture, and this magnitude is optimal to measure galaxy colours that are unbiased relative to the S/N of the image. SDSS also provides the *dereddened magnitude*, which is effectively the extinction-corrected model magnitude, which will be the magnitude used in this work for photo-z estimation.

Another important magnitude fit is the *Petrosian magnitude* (Petrosian 1976). While model magnitudes for each galaxy suffer from having different profiles with different edges, the Petrosian magnitude measures fluxes in apertures determined by the shape of its profile and not its amplitude: it is measured with a constant fraction of total light with an aperture defined by the *Petrosian ratio* $\mathcal{R}_{\text{P}}(r)$ independent of the position and distance of the the object. Here we invoke the SDSS definition of the Petrosian ratio (Blanton et al. 2001), which is the ratio of flux in an annulus $0.8r - 1.25r$ to the flux within r:

$$\mathcal{R}_{\text{P}}(r) = \frac{r^2}{(1.25r)^2 - (0.8r)^2} \frac{2\pi \int_{0.8r}^{1.25r} I(x)\, x\, dx}{2\pi \int_0^r I(x)\, x\, dx}. \tag{55}$$

$\mathcal{R}_{\text{P}}$ is evaluated at $r = r_{\text{P}}$ where r_{P} is the *Petrosian radius*, and SDSS defines r_{P} to satisfy the equation $\mathcal{R}_{\text{P}}(r_{\text{P}}) = 0.2$ (a matter of choice, see Strauss et al. 2002, for more details). This defines the Petrosian radius r_{P} as the radius at which the flux in an annulus about r_{P} is one-fifth of the mean flux within r_{P}. r_{P50} and r_{P90} are also defined in a similar way, replacing 0.2 with 0.5 and 0.9, and will be used to define the concentration index in Section 2.2.2. The value of r_{P} derived from Equation 55 may have 2 solutions (in which case the outer solution is selected), and sometimes may even yield no solution due to the noise, substructure and the finite object size associated with the definition, this happens especially for faint objects with r-band $r > 20.5$ in SDSS. The Petrosian magnitude is then calculated based on the flux within a circular aperture of radius $2r_{\text{P}}$

(also a matter of choice), which is known as the *Petrosian aperture*. Therefore the Petrosian flux can be obtained via the equation below:

$$F_P = 2\pi \int_0^{2r_P} I(x)x\,dx. \tag{56}$$

The Petrosian aperture $2r_P$ is set by the profile of the galaxy in the r-band, and the same aperture is used for the other bands for consistent colour measures. The Petrosian ratio is insensitive to inclination since $2r_P$ is defined to be large enough to contain nearly all of the flux of a galaxy yet keeping the sky noise at minimum (Blanton et al. 2005). In the absence of noise, $2r_P$ is found to be unaffected by foreground extinction, cosmological dimming of surface brightness, sky brightness, and seeing. In this work Petrosian magnitudes are only used for target selection of the MGS.

Lastly we also define the *Kron magnitude*, which is not used in SDSS but instead by CS82. Kron (1980) defined the *Kron radius* r_K, a luminosity-weighted radius (also known as the first moment of an image),

$$r_K(r) = \frac{2\pi \int_0^r I(x)x^2\,dx}{2\pi \int_0^r I(x)x\,dx}. \tag{57}$$

r_K is evaluated at $r_K(r_{s1})$, where r_{s1} is the radius where the integrated flux contains 1% sky flux, in which Kron argued that a radius of $2r_K(r_{s1})$ would cover more than 90% of an object's total flux. The modern definition of the Kron magnitude uses $2.5r_K(r_{s1})$ instead, which is also the value used by SExtractor (Bertin and Arnouts 1996), which named this magnitude as `mag_auto`. While the Kron magnitudes may seem to have been defined such as to minimise the sky flux, it has been reported in some instances to be missing up to half of a galaxy's light, especially if the intensity is fitted with a Sérsic profile with large n (Graham and Driver 2005).

We now summarise the uses of different magnitudes in this work (the samples and surveys listed below will be described in Chapter 4). Photometry obtained from SDSS are all dereddened model magnitudes; the CS82 i-band is a Kron magnitude; the COSMOS $uBVriz$ magnitudes are fixed-aperture magnitudes; while the PAU narrow band magnitudes are measured by a fixed percentage of light (value not finalised at the time of writing). The CS82 PSF magnitude is used only in the production of the CS82 morpho-z catalogue as a proxy for star-galaxy separation. While other magnitudes were not directly used, morphology derived from their particular fits were used: radii, axis-ratio and shape parameters were derived from de Vaucouleurs and exponential profiles, the Petrosian radii were used to calculate the concentration index, and the Sérsic index from the Sérsic profile was used too. All these morphological parameters will be properly defined in Section 2.2.2.

As this work uses photometry and spectroscopy from many sources, we do not make an exhaustive effort to specify the type of magnitude every time a magnitude is mentioned (especially when discussing survey-specific target selection magnitude cuts). However in the cases when the mention of a specific type of magnitude matter, we will use the following subscripts to differentiate the magnitudes: fib (spectroscopic 3″ fibre aperture, SDSS only), PSF (PSF), deV (de Vaucouleurs), exp (exponential) Ser (Sérsic), mod (model), cmod (composite model), Pet (Petrosian) and auto (Kron).

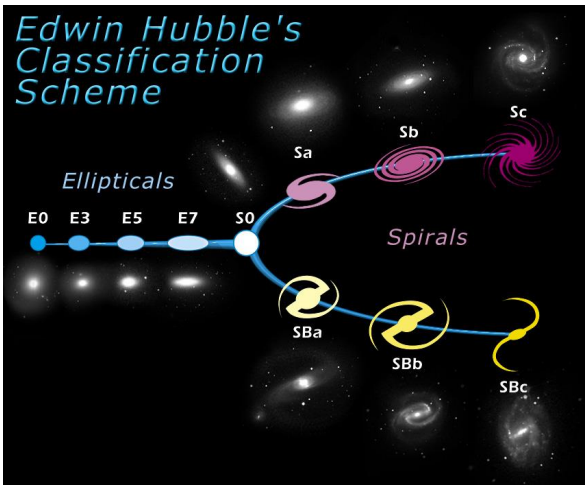

Figure 4: The Hubble Sequence, a morphological classification scheme for galaxies. Figure taken from https://www.spacetelescope.org.

2.2 MORPHOLOGY

Galaxy morphology refers to the structural properties of galaxies, which includes disks, bulges, bars and rings. Galaxies are commonly classified into two categories: *elliptical galaxies*, which have approximately ellipsoidal shape and smooth featureless brightness profiles; and *disc galaxies*, which have a prominent disc-like structure, and could be further classified based on the presence of bars, spiral shapes or bulges. Galaxy morphology helps us to understand the formation history and evolution of the galaxy. For example, elliptical galaxies are generally thought to have formed early in the history of the Universe, look redder and composed of older low mass stars, while disc galaxies look bluer and and have active on-going star formation. Galaxies that fall in between are classified as *lenticular galaxies*; galaxies that do not have a distinct shape are classified as *irregular galaxies*; while galaxies that have unusual properties due to tidal interaction with other galaxies are classified as *peculiar galaxies*. The most common galaxy morphological classification system can be visualised using the *Hubble Sequence* (Hubble 1926) in Figure 4. In the diagram, E refers to elliptical galaxies, with numbers indicating its ellipticity; S refers to spiral galaxies, further subcatogerised by those with bars (SB) or without (Sa, Sb and Sc). There are in fact other morphological classification systems (de Vaucouleurs, Yerkes etc) and could be further refined with the help of machine learning (Lahav et al. 1996; Banerji et al. 2010), but these shall not be discussed further as they are beyond the scope of this work.

For the purpose of this thesis, we will not dive into an in-depth discussion on specific galaxy morphological structures and their connections with galaxy formation and evolution. However, we are more interested in how certain morphological structures (which could be represented by numerical values) would be able to influence the results of photo-z estimation. These galaxy morphological parameters like radius, axis-ratio and ellipticity (full list discussed in Section 2.2.2) are estimated during the model fitting process, and technically 'come for free' with the fluxes obtained from photometric surveys.

2.2.1 *Star-Galaxy Separation*

Before we proceed to list the different morphological parameters used in this work, it is worth clarifying our definition of star-galaxy separation. Firstly, it is worth noting that star-galaxy separation can be done photometrically, morphologically and spectroscopically. Objects can be classified as *extended objects* or *point sources* photometrically based on the difference in magnitude fits used. For example, objects are classified as extended objects (`TYPE=3`) in SDSS if they satisfy the equation $r_{\mathrm{PSF}} - r_{\mathrm{cmod}} > 0.145$ (and `TYPE=6` otherwise), where r_{PSF} and r_{cmod} are the r-band magnitude of the PSF and cmodel fits respectively (Stoughton et al. 2002), essentially selecting objects which have higher brightness in a cmodel fit than a PSF fit. On the other hand, objects are classified spectroscopically as `CLASS=STAR`, `CLASS=GALAXY` and `CLASS=QSO` based on their spectra.

From this, it is important to note that an object with `TYPE=3` does not necessarily mean it has `CLASS=GALAXY`, and vice versa. The photometric classification of `TYPE=3` and `TYPE=6` objects by SDSS is obviously imperfect and probabilistic, and thus other star-galaxy separation algorithms have been introduced by many, e.g. using infrared magnitudes (Bundy et al. 2015), machine learning (Soumagnac et al. 2015; Kim and Brunner 2017) or quasar variability (Peters et al. 2015). In this work, we treat the spectroscopic classification as a proxy for a 'true' star-galaxy separator when in comparison with the photometric classification.

Secondly, we do not use spectroscopic redshifts or photometry of stars at all in this work. Thus, in our context, 'star-galaxy separation' actually means 'quasar-galaxy separation', or rather, 'point source-extended object separation'. We do this by ensuring that our training or testing samples do not have (1) objects with `CLASS=STAR` or have redshift $z = 0$, and (2) objects with `TYPE=6`. In other words, we are essentially only dealing with objects with `TYPE=3` and `CLASS=GALAXY` (or simply 'extended galaxies').

In certain sections of this work, quasi-stellar objects (QSOs, or quasars) are also used. A quasar is a type of active galactic nucleus (AGN), which is a compact region at the centre of a galaxy that has a much higher than normal luminosity. Since quasars are very luminous objects at very high redshifts, they are usually observed as point sources like stars. Quasars are easily differentiated spectroscopically from stars and galaxies due to their distinct spectral features (discussed in Section 2.3), however it is usually not easy to separate them from galaxies and stars photometrically (see Peters et al. 2015, for a good discussion on this topic). In a way, objects with `TYPE=3` and `CLASS=QSO` are known as 'extended quasars', or quasars with visible bulges of the host galaxies which we expect would contaminate a sample of galaxies by a non-negligible amount. The impact of the inclusion of these objects photo-z estimation is evaluated in Section 6.4.2.

2.2.2 *Morphological Parameters*

In the following paragraphs, the morphological parameters used in this work will be properly defined, and the motivation to use them to improve photo-z's are discussed. A summary of the morphological parameters and their symbols are shown in Table 1. When different model fits are used, subscripts deV (de Vaucouleurs) and exp (exponential) are denoted to distinguish between them.

Table 1: List of morphological parameters used in this work along with their symbols.

Symbol	Name
r	Radius
s	Circularised radius ($\sqrt{\text{size}}$)
q	Axis-ratio
μ	Mean surface brightness
n	Sérsic index
C	Concentration index
p	Shape probability
f	`fracDeV`
Q, U	Stoke's Q and U parameters

First we define the *radius* r of a galaxy, measured in arcseconds. This is equivalent to the effective angular radius r_e obtained after fitting a model to the galaxy, it is also the semi-major axis of the galaxy. We then have the *axis-ratio* $q = \frac{r_b}{r_a}$, which is the ratio between the semi-minor axis r_b and the semi-major axis r_a of the galaxy. This parameter is used as a main measure for the object's ellipticity e, where q is related to e via the following equation:

$$e = \sqrt{1 - \frac{r_b^2}{r_a^2}} = \sqrt{1 - q^2}. \tag{58}$$

q is not the only way to measure ellipticity, we can also measure it using *Stokes' Q and U parameters* provided by SDSS,

$$Q = \frac{r_a - r_b}{r_a + r_b} \cos 2\theta, \tag{59}$$

$$U = \frac{r_a - r_b}{r_a + r_b} \sin 2\theta. \tag{60}$$

where θ is the position angle of the galaxy. Similar to q, Q and U can be related to the ellipticity via the equation

$$e = \sqrt{1 - \left(\frac{1 - \sqrt{Q^2 + U^2}}{1 + \sqrt{Q^2 + U^2}}\right)^2}. \tag{61}$$

The advantage of Q and U to q is that they are model independent as r_a and r_b here are measured with respect to a fixed isophote, however SDSS has reported that their performances are not ideal at low S/N, and later replaced them with adaptive moments m_{e1} and m_{e2} (Bernstein and Jarvis 2002). Q and U were used as motivated by the work of (Way and Srivastava 2006), however in most parts of this work, q will be used as the main measure of ellipticity. We attempt to see if the ellipticity of galaxies affects the photo-z estimation.

With the axis-ratio defined, we can now introduce the *circularised radius* s, where

$$s = r\sqrt{q} = \sqrt{r_a r_b}, \tag{62}$$

which is a form of measure for galaxy size (or area) that is independent of the galaxy's ellipticity. At $z < 1$ we expect galaxies which are further away to have smaller angular radius, therefore we want to test the significance of s in photo-z estimation. In Section 6.5 we even used two different circularised radii for comparison, one from SDSS ($s_{\rm SDSS}$) and the other from CS82 ($s_{\rm CS82}$) to study if the quality of morphology affects the photo-z results.

The *mean surface brightness* μ of an object is the brightness per unit solid angle as seen in projection on the sky, measured in units of mag arcsec^{-2}. μ is a derived morphological quantity, having the equation below:

$$\mu = m - A + 2.5 \log_{10}(2\pi s^2), \tag{63}$$

where m is the apparent magnitude and A the extinction. μ is in fact analogous to photometric luminance, it should be constant with distance in a static universe since the light received and the apparent area of an object both drop proportionally to the square of its distance. However, this is not the case in an expanding universe, the contributions of redshift cause a surface brightness dimming proportional to $\log(1+z)^4$ (Tolman 1930), which causes a significant bias in detection of galaxies at high redshift (Calvi et al. 2014). Therefore it is natural to expect μ be able to improve photo-z estimations.

Other than size and ellipticity, we are also interested to know if the type or shape of galaxy (i.e. elliptical or disc) affects the photo-z estimation, and a few morphological parameters listed below are indicators of shape and type. First we have the *Sérsic index* n, as introduced in Section 2.1.6 which controls the slope of the Sérsic profile. A value of $n = 1$ gives an exponential profile, while $n = 4$ gives a de Vaucouleurs profile, and we intend to test the dependence of redshift with n. As the SDSS pipeline does not include Sérsic fits, n is obtained from CS82 and the New York University Value-Added Galaxy Catalogue (NYU-VAGC, Blanton et al. 2005) and later cross-matched with SDSS photometry.

The *concentration index* C indicates how centrally concentrated a galaxy's stellar distribution is. There are many different definitions of concentration (see Graham and Driver 2005, for a full list of definitions), but in this work we use the SDSS definition,

$$C = \frac{r_{\rm P90}}{r_{\rm P50}}, \tag{64}$$

where $r_{\rm P90}$ and $r_{\rm P50}$ are the radii containing 90% and 50% of the Petrosian flux, respectively. Galaxies with de Vaucouleurs profile have $C \sim 3.33$ while galaxies with exponential profiles have $C \sim 2.32$ (Strateva et al. 2001). The values of C obtained from SDSS however are not corrected for seeing, and more than 40% of galaxies in fainter samples do not have defined values of C due to the definition of Petrosian magnitudes, therefore C is not used in the later parts of Chapter 6.

f (or `fracDeV` in the SDSS) as mentioned in Section 2.1.6 is the fraction of the galaxy fitting the de Vaucouleurs profile as compared to the exponential profile[2]. The usage of this parameter was also motivated by the work of Way and Srivastava (2006) and Vincent and Ryden (2005), where the latter has suggested that it is a good proxy for the Sérsic index.

2 This was earlier defined as $f_{\rm deV}$, we have dropped the subscript for consistency with other morphological parameters.

Finally, we define a new parameter p called the *shape probability*, which indicates if the object's shape is closer to a disc galaxy (exponential fit) or to an elliptical galaxy (de Vaucouleurs fit):

$$p = \frac{\chi^2_{\text{deV}}}{\chi^2_{\text{deV}} + \chi^2_{\text{exp}}}. \tag{65}$$

p takes values between 0 and 1, it compares the reduced χ^2 values of both fits, therefore the higher the value the more likely it is a disc galaxy. This is motivated by the observation that there are more disc galaxies than elliptical galaxies in a complete sample at $z = 0$, and therefore we intend to see if adding a shape parameter would improve the photometric redshift of these galaxies.

2.3 SPECTROSCOPY

In Chapter 1 we have discussed the prominent methods to determine distances to distant galaxies including standard candles, standard rulers and empirical relationships. We have also defined redshift as the shifting of light towards the redder spectrum. Section 2.1 further assists us to continue our discussion on the distance ladder which we left off earlier. In this section we will discuss *spectroscopy*, the technique of splitting electromagnetic radiation into its constituent wavelengths (or *spectra*), which is an important method to obtain redshifts of galaxies and quasars which are very far away from us.

In Chapter 1, we have not mentioned how people know that the light from celestial bodies observed has been redshifted, and redshifted with respect to what. In *astronomical spectroscopy*, one would look at features in the redshifted galaxy spectra, and compare these with known rest-frame spectra of molecules and atoms on Earth. Fluxes of celestial objects are measured in a distribution of wavelengths, it is recorded as a spectral energy distributions (SED) and the shift in wavelength can be determined. An example of a rest-frame galaxy spectrum is shown in Figure 5.

The first astronomical spectral lines discovered were the dark absorption lines in the Sun, they were observed by William Wollaston (1802) and later characterised by Joseph von Fraunhofer (1814, thus getting the name 'Fraunhofer lines'), and even later identified as elements in the solar atmosphere by Gustav Kirchhoff (1860). Therefore spectroscopy not only allows us to study distances to galaxies, they also help us to determine the chemical composition and physical properties like the temperature, density and mass of galaxies. If different parts of the galaxy have different line-of-sight velocities due to rotation, this would result in a Doppler broadening of the spectral lines, in which the line width would be proportional to the velocity dispersion of the galaxy.

There are two kinds of spectral lines observed in a spectrum. When a gas element is heated to high temperatures, *emission lines* are radiated from the gas. These emission lines are most often found in star-forming regions of the galaxy, centred at certain wavelengths. They are photons emitted due to excited electrons making a transition from a high energy state to a lower energy state. The photon energy E_γ is the energy difference between the two states, the energy of the photon can be represented by the equation $E_\gamma = hf$, where h is the Planck constant and f is the frequency of the photon. Since each element has unique energy level differences, these lines are

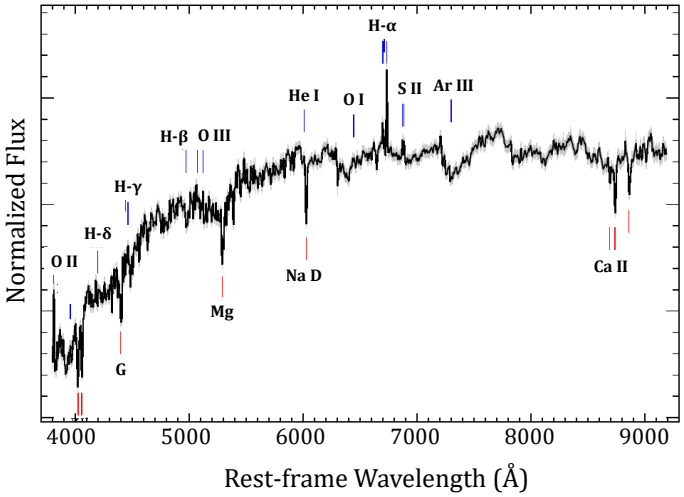

Figure 5: A rest-frame spectra of an LRG obtained from the SDSS Legacy Survey, with the respective emission lines labelled. Figure obtained and modified from http://skyserver.sdss.org.

unique to each element, and therefore once we see these signature lines we can determine the element compositions in the gas.

Absorption lines follow the same principle, however they occur when photons are absorbed by a colder source in front of it, leaving a trough in the spectra. In the context of galaxies, radiation from the centre of the galaxy are absorbed by atmospheres of large stars, gas clouds or interstellar medium. Radiation is more likely to be absorbed at frequencies with similar energy differences between two quantum mechanical states of the atoms, therefore absorption lines could help us identify compositions of interstellar clouds.

2.3.1 *Hydrogen Spectra*

Since hydrogen is the most abundant element in the universe, spectra originating from the hydrogen atom are commonly observed. The hydrogen atom has only one electron around its nucleus, thus the wavelength λ of the photon emitted when its electron jumps from level n_2 to n_1 can be modelled using the *Rydberg formula*:

$$\frac{1}{\lambda} = R_H \left(\frac{1}{n_1^2} - \frac{1}{n_2^2} \right) \tag{66}$$

where $R_H = 1.097 \times 10^7 \text{ m}^{-1}$ is the *Rydberg constant*. Therefore with the wavelength of emission line measured we will be able to know the transition levels of the electron. Below we summarise a few common hydrogen emission lines which play a big role in astronomical spectroscopy.

The *Lyman-α line* is observed in the electron transition from energy levels 2 to 1. It has a wavelength of 1215.7 Å which lies in the ultraviolet region. The detection of the Lyman-α line is usually carried out by satellite-borne instruments, except for distant sources where it has been redshifted to the visible spectrum. In fact, the Lyman-α line splits into a doublet due to fine structure perturbation but is rarely resolved in astronomical spectroscopy. Lyman-α lines are

prominently observed in the *Lyman-α forests*. They are a series of absorption lines in the spectra of very distant galaxies and quasars, absorbed by intergalactic gas in between. The Lyman-α forest is an important probe of the intergalactic medium, it can be used to determine the density and temperature of clouds containing neutral hydrogen. Quasars are important 'backlights' which illuminate neutral hydrogen through Lyman-α absorption, this allows us to map large-scale structure in the foreground of each quasar.

The *H-α line* is observed due to the electron transition from energy levels 3 to 2, in fact it is part of a series of *Balmer lines* from electron transitions from an arbitrary level to level 2, named systematically H-α, H-β, H-γ, H-δ etc in ascending order of energy level. H-α has a visible wavelength of 6563 Å, it is an important line used to detect the presence of hydrogen. H-α produces a bright red line in nebula or H II regions, which in fact is what makes the Orion nebula look pink, and could indicate the shape and extent of a nebula cloud. The H-α line and other Balmer lines are prominently seen in the solar spectrum and are mainly found in spiral galaxies (especially in the spiral arms) and also irregular galaxies. Since H-α occurs very often when electrons recombine with ionised hydrogen, it is also used to trace ionised hydrogen in gas clouds.

Another important feature of the Balmer series is the *Balmer break*, in which the spectra show a lack of radiation with wavelength shorter than 3646 Å, corresponding to the electron completely ionised from energy level 2 (the Balmer limit). Galaxies where this continuum absorption are observed are known as *Balmer-break galaxies*, where this spectral feature is common in Type A stars, which suggests that Balmer-break galaxies are star-forming galaxies. The Balmer break allows high redshift star-forming galaxies easily be identified photometrically, they can be selected via simple colour-colour diagrams (Díaz Tello et al. 2016).

Its analogous partner is the *Lyman-break galaxy*, also a star-forming galaxy, showing a lack of radiation with wavelength shorter than 912 Å, corresponding to electrons ionised from the first energy level (the Lyman limit). This radiation is absorbed by neutral hydrogen, and is typically observed in the optical at redshift $z \sim 3$, but possible at lower redshifts with ultraviolet photometry. They are easily identified photometrically too since the break caused images of these galaxies to be undetected in certain filters. The strength of the continuum absorption of both breaks depends on the temperature and density of the region responsible for the absorption, and is a proxy for the age of the stellar population. An example of the Balmer break absorption continuum is shown in Figure 6.

Finally we also have the 21 *cm line* (or hydrogen line / H I line), which is a hyperfine transition in the hydrogen ground state in neutral hydrogen atoms. As its name suggests, it is observed far in the radio region, it is in fact a highly forbidden process with an extremely small transition rate, yet commonly observed with little interference in hydrogen clouds (H I regions) because radio waves can penetrate large clouds of interstellar dust that are opaque to visible light. However ground based measurements are affected by the interferences from television transmitters and the ionosphere, therefore it must be measured from very secluded sites, or on the far side of the moon. The 21 cm line was first detected by Ewen and Purcell (1951), its discovery eventually allowed the mapping of a neutral hydrogen map of our universe. The Milky Way was in fact determined to be a spiral galaxy when the Doppler shift in neutral hydrogen is seen in the spiral arms with the help of the 21 cm line (van de Hulst et al. 1954). The 21 cm line is currently the only way for us to probe the dark ages: from recombination to reionisation, it provides a picture

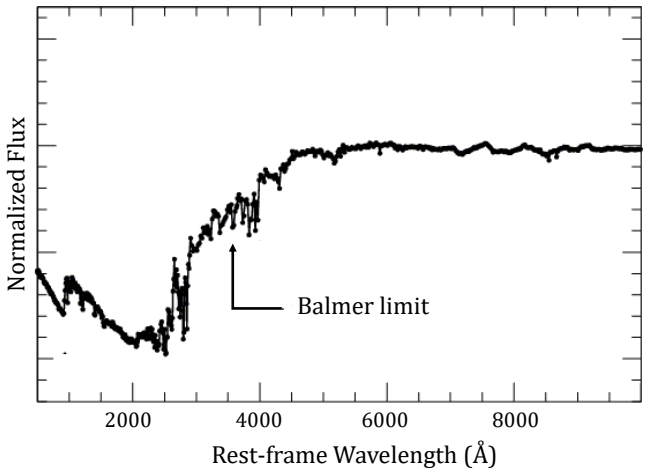

Figure 6: An example of the Balmer break absorption continuum to the left of the Balmer limit. Figure from Massarotti et al. (2001).

of how our universe was ionised, since neutral hydrogen which has been ionised by radiation will appear as holes in the 21 cm background.

2.3.2 *Galaxy Spectra*

Knowing that galaxies are made of a combination of million of stars, we expect that galaxy spectra should be a combination of different types of stellar spectra. The Lyman-α, H-α, 21 cm line and the continuum breaks mentioned in Section 2.3.1 are all observed in galaxies, and the different features can help us identify different types of galaxies at different redshifts.

There are a few other important non-hydrogen spectral features which further help us characterise galaxies. For example the 4000 *Å break*, which is seen prominently in LRGs, this break is caused by absorption of high energy radiation by metals in the stellar atmospheres and the lack of hot blue stars. Therefore just as how H-α helps us identify spiral galaxies, the 4000 Å break helps us identify LRGs (which are primarily elliptical galaxies) in the optical at the redshift range of $0.12 < z < 0.55$.

Most nearby and low-redshift galaxies are identified via prominent absorption lines, as less than 10% of these galaxies have strong emission lines (Bolton et al. 2012). Some major absorption lines observed are the Mg b triplet (5180 Å), Ca line (5270 Å), and the Na I doublet or D lines (5890 Å). Galaxies which are further rely on some narrow-peaked emission lines for identification, these bluer galaxies are also known as *emission line galaxies* (ELGs). Other than the H-α line, other major emission lines include the O II doublet (3727 Å and 3729 Å) which is well resolved at $z \sim 0.95$ (Drinkwater et al. 2010), and the O III lines (4959 Å and 5007 Å) which are observable in the optical from $0.25 < z < 0.68$.

Emission lines are also important for detecting AGNs. An AGN is a compact region at the centre of a galaxy that has much higher luminosity than usual, caused by the accretion of matter by a supermassive black hole. The emission lines of AGNs are broadened due to the emitting material revolving around the black hole with high speeds, causing Doppler broadening. AGNs can be identified by measuring the amplitude ratios of its composition lines (Bolton et al. 2012).

A quasar is in fact a type of AGN. Quasar spectra can be identified by their bright blue continuum and broad lines. Therefore we see that emission and absorption lines are very important features to identify very distant objects, allowing redshifts to be measured even beyond the *redshift desert* of $1.5 < z < 2.5$, a region where ground-based observations in the optical and ultraviolet have been proved very difficult in the past (Renzini and Daddi 2009). However, quasars are difficult to identify photometrically as they are usually point sources which have to be separated from stars. Stars have roughly blackbody spectra, while quasars have spectra that are characterised by a featureless blue continuum and strong emission lines. With these distinctive features, SDSS was able to define a stellar locus in a colour-colour diagram, and quasars would lie in outlier regions which allow them to be separated from stars (Richards et al. 2002).

2.3.3 *Spectrograph*

In astronomy, the instrument used to obtain galaxy spectra is called a *spectrograph*. The spectrograph is placed on a telescope where light is focused into and passes through a slit mask (to mask light from other sources), the collimated light is then dispersed by the spectrograph so that its spectra can be analysed.

Many types of dispersing elements have been used in spectrographs. In earlier years prisms were used, however due to the large physical size of prisms, currently most spectrographs use either reflecting or transmission diffraction gratings. A grating is a piece of glass with periodic rulings on its surface enabling it to diffract light. Polychromatic light entering a diffraction grating would be dispersed and produce multiple orders of spectra, and the diffracted beams corresponding to consecutive orders may overlap, depending on the spectral content of the incident beam and the grating density.

Gratings are often blazed in order to concentrate light away from zero order to higher orders for better resolution. A *blazed grating* has grating lines which have triangular, sawtooth-shaped cross section, light is passed through at a certain angle such that maximum optical power is concentrated in a desired diffraction order for a certain wavelength, while power at other orders minimised. The specific angle and wavelength are called the *blaze angle* and *blaze wavelength* respectively, therefore blazed gratings are optimised based on these parameters. Other than gratings, a grism (grating on a prism) is also a popular choice of dispersing element. A grism only allows light of a certain wavelength range to pass through, this is advantageous for a single camera that is used for both photometry and spectroscopy.

The dispersed light is then captured by the CCD and recorded. Since the dispersed light is now a small fraction of the total light from the object, long integration times are required for the CCDs to obtain a sufficiently high $^{S}/_{N}$. The obtained spectra will then be processed: the spectra is extinction corrected, the sky spectrum will be subtracted from the flux, and the flux will be calibrated with star spectra. The spectra are calibrated by comparing to the spectra of gas lamps, like a helium-neon-argon lamp.

Modern spectrographs are able to do *multi-object spectroscopy*, obtaining spectra for many objects at once to improve its efficiency. The simplest way is in fact to use *multi-slit spectrographs*, which requires a custom-made cut of short slits at the position of each target for each pointing. This method allows total control on slit width and length and good sky subtraction, however

it is limited by the number slits that could be cut before overlapping of target occurs. Another method is to use *fibre-fed multi-object spectrographs*, which is to place a fibre at each target position at each pointing. This allows a uniform wavelength coverage and an efficient use of detector area, however the fibre receives both sky and target light, which requires proper sky subtraction, and the finite fibre size would also cause fibre collision (see Guo et al. 2012, for an interesting discussion and solution on this issue).

Another important type of spectrograph is the *integral field spectrograph*, which is a spectrograph equipped with an integral field unit (IFU) that allows one to gather spectra over a two-dimensional field-of-view. In other words, an IFU is able to obtain spectra for each part of an extended object like galaxies, and this is helpful to obtain rotational properties and velocity dispersions of galaxies. A notable spectrograph that has an IFU is the Visible Multi-Object Spectrograph (VIMOS), which will be discussed in Section 4.3.5.

The performance of a spectrograph can be measured via its spectral resolution R_S. It is a measure of its ability to resolve features in the electromagnetic spectrum:

$$R_S = \frac{\lambda}{\Delta\lambda}, \tag{67}$$

where λ is the wavelength of light received and $\Delta\lambda$ is the smallest wavelength resolvable. Typical modern spectrographs for large galaxy surveys have $R_S \sim 1000$. R_S can be improved based on slit width or grating choice.

On the subject of spectrographs, perhaps it is worth mentioning the first modern spectroscopic survey, the *Harvard-Smithsonian Centre for Astrophysics* (CfA) Redshift Survey which began observing in 1977 (Tonry and Davis 1979). The CfA survey aimed to measure the properties of large-scale structures of the Universe. It discovered the large-scale structure *Great Wall*, and it provided the first evidence that galaxies lie on web-like structures with clusters of galaxies at their junctions (Geller and Huchra 1989). CfA used a photon-counting Reticon detector (Davis and Latham 1979) on the Mount Hopkins 60″ telescope, observing as deep as $m_B \sim 14.5$. It used a 600 line/mm grating which covered wavelengths between 4300 Å and 6900 Å at a resolution of 4.8 Å, and the data were recorded on floppy disks. It was able to observe 20 galaxies per night, and as we can see from the paragraphs below, the development of fibre-optic and multi-slit spectrographs after the 1990s sped up and improved the quality of measurements.

It is worth noting that most large spectroscopic surveys are ground based, and the observation wavelength range is limited by the Earth's atmosphere. X-ray and ultraviolet spectroscopy require the use of satellite telescopes since ozone and oxygen molecules absorb light with wavelengths under 3000 Å. Infrared wavelengths on the other hand are absorbed by H_2O and CO_2 in the atmosphere. Infrared telescopes thus have to be built at high and dry places, while radio signals require large antennas or radio dishes to obtain a high enough S/N. The limits of ground-based spectroscopic surveys can be visualised in Figure 7. Sometimes spectroscopic surveys are limited by the sensitivity of their CCDs too, for example the silicon CCDs has a sensitivity limit of 11 000 Å (Fukugita et al. 1996).

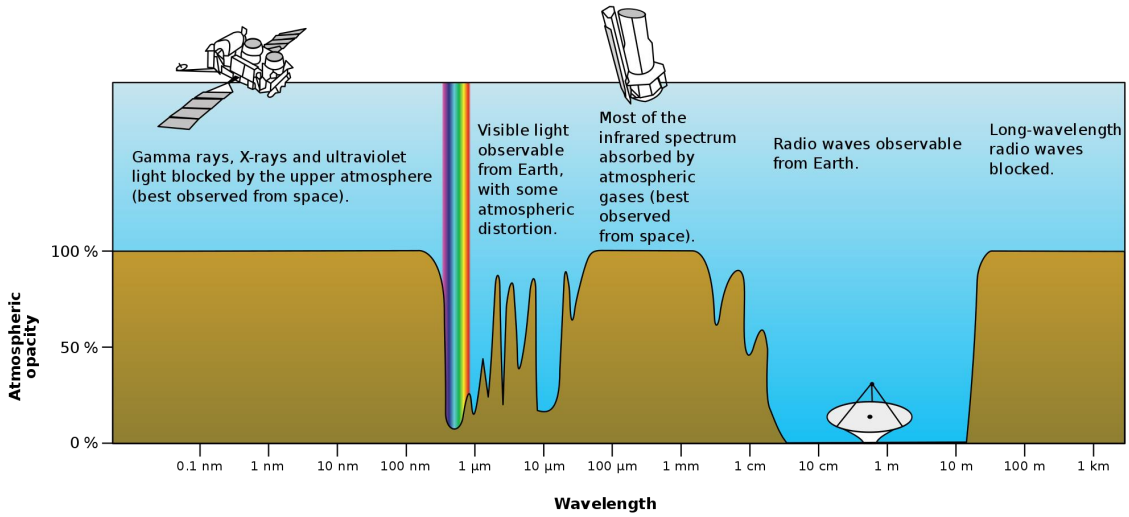

Figure 7: Visualisation of the limits of ground-based spectroscopy. Figure obtained from https://commons.wikimedia.org.

3

PHOTOMETRIC REDSHIFTS: PAST AND PRESENT

"I am the Alpha and the Omega," says the Lord God, "who is, and who was, and who is to come, the Almighty."

Revelations 1:8

3.1 INTRODUCTION

In the previous chapter we have studied the principles behind spectroscopic redshift estimation. The collective efforts of these surveys add the radial dimension to the sky map: a 3 dimensional sky map holds the key to fundamentally all kinds of cosmological research, from galaxy formation to large-scale structure distributions, pushing our observations further into the past to understand how our Universe came to be today. Spectroscopic redshifts (spec-z's or $z_{\rm spec}$ for short) of individual galaxies and quasars are important in their own right, however for recent cosmological research like weak lensing, clustering and intrinsic alignment, redshifts are needed in large volumes to produce precise and accurate galaxy redshift distributions $n(z)$.

The volume of spec-z's at the present time however is severely insufficient for many current cosmological surveys, partly because (1) unlike photometry, spectroscopy is a very time-consuming and expensive process; and (2) the accuracy of spectroscopic measurements is always limited by the technology and optics at that time. Therefore in order to produce redshifts for ideally all objects in large galaxy samples, the concept of *photometric redshifts* was born. Photometric redshifts (photo-z's or $z_{\rm phot}$ for short) use the photometry of multiple broad band filters to produce redshifts, or in very simple words, they are technically 'redshifts from very low resolution SEDs'.

In the 1990s, Koo (1999) attempted to give a proper definition for photo-z by defining it to be a redshift derived from only images or photometry with spectral resolution of $R_{\rm S} < 20$ to exclude redshift estimation methods derived from slit spectra, narrow band images and spectrometers. However in the present day, this definition is not followed, since redshifts produced from narrow band surveys like PAU with $R_{\rm S} \sim 50$ are considered photo-z's, while redshifts produced from a prism by the Prism Multi-object Survey (PRIMUS, Coil et al. 2011) with $R_{\rm S} \sim 40$ are considered spec-z's. No formal modern definition has been given, but photo-z's can generally be understood as 'redshifts estimated using photometry through bandpass filters, without the use of spectra'.

Photo-z's are widely sought after especially for galaxy evolution and cosmology surveys. Early needs for photo-z's include the search of primeval galaxies (Partridge 1974; Madau et al. 1996;

Table 2: List of photo-z requirements for major current and upcoming large sky surveys. Here σ_{68} is the 68th percentile error, η the outlier fraction and $\sigma_z = \sigma_{\mathrm{RMS}}$ the scatter (or root-mean-square error).

Survey	**Source**	**Photo-z Requirement**
DES	Abbott et al. (2005)	$\sigma_{68} < 0.12$, $\eta_{2\sigma} < 0.1$, $\eta_{3\sigma} < 0.015$
Euclid	Laureijs et al. (2011)	$\sigma_z < 0.05(1+z)$, $\eta_{0.15(1+z)} < 0.1$, bias$< 0.002(1+z)$
HSC	Aihara et al. (2018)	-
KiDS	de Jong et al. (2013)	$\sigma_{\mathrm{RMS}} < 0.03(1+z)$
LSST	Ivezić et al. (2008)	$\sigma_z < 0.02(1+z)$, $\eta_{3\sigma} < 0.1$, bias$< 0.003(1+z)$
PAU	Castander et al. (2012)	$\sigma_{68} < 0.0035(1+z)$
WFIRST	Spergel et al. (2013)	-

Steidel et al. 1996) and radio galaxies (Puschell et al. 1982) at redshift $z > 3$. Later applications of photo-z's include the study of the galaxy luminosity function (SubbaRao et al. 1996; Ramos et al. 2011), weak lensing (Mandelbaum et al. 2005; Benjamin et al. 2013; Kuijken et al. 2015), galaxy clustering (Myers et al. 2009; Soltan and Chodorowski 2015; Jouvel et al. 2017), neutrino mass (Thomas et al. 2010), galaxy group finding (Gillis and Hudson 2011; Castignani and Benoist 2016) and the search for high redshift galaxy clusters (Castignani et al. 2014). Recent applications of photo-z's include follow-ups on gravitational wave detections (Antolini and Heyl 2016) by the Laser Interferometer Gravitational-wave Observatory (LIGO, Barish and Weiss 2008) and structure growth studies (Bianchini and Reichardt 2018).

In order to fulfil the current redshift needs in cosmology surveys, many current and upcoming surveys such as the Hyper Suprime-Cam (HSC, Aihara et al. 2018), Kilo-Degree Survey (KiDS, de Jong et al. 2013), Dark Energy Survey (DES, Abbott et al. 2005), LSST (Ivezić et al. 2008), Wide Field Infrared Survey Telescope (WFIRST, Spergel et al. 2013) and Euclid (Laureijs et al. 2011) have set photo-z requirements to ensure that they meet current research demands and science goals. A summary of photo-z requirements set by the surveys mentioned above are listed in Table 2.

In this chapter we will discuss the history of photo-z's: the different estimation methods, and finally the current efforts on improving them, which leads us to the ultimate aim of this research.

3.2 HISTORY OF PHOTOMETRIC REDSHIFTS

Photo-z's have been studied as early as the 1960s, it started off when the scientific community placed interest in a redshift-magnitude relation. Baum (1962) was the first to develop a technique for measuring redshifts photometrically. In a paper entitled *Photoelectric Magnitudes and Redshifts*, he used a photometer and 9 bandpasses to compare the displacement in spectral energy distributions (SEDs) between galaxies in the Virgo cluster and Cl0925 + 2044 to predict redshifts of other clusters up to redshift $z \sim 0.46$. With his results, a redshift-magnitude relationship was plotted for 8 galaxy clusters, and this was possible due to the 4000 Å break feature which is most prominent on elliptical galaxies.

Later, a technique which uses photometry to constrain redshifts is devised by Partridge (1974), known as the *Lyman-drop technique* (or ultraviolet drop-out technique). This technique uses the sudden drop in flux in the rest-frame ultraviolet filters due to the Lyman break to predict redshift

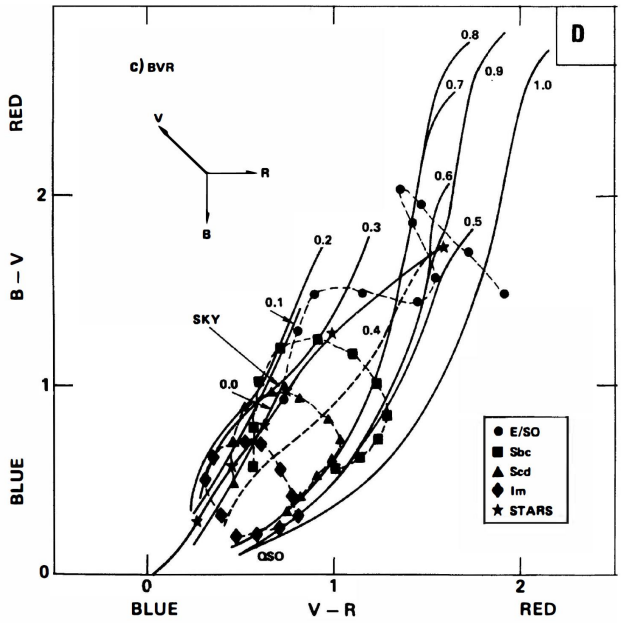

Figure 8: Original figure from Koo (1985) which shows the iso-redshift lines for galaxies of different spectral types plotted on a colour-colour diagram.

ranges of galaxies between redshift $2.4 < z < 3.4$. Although this was technically not a photo-z technique since it was only used to discern very high redshift galaxies, his work however motivated Madau et al. (1996) and Steidel et al. (1996) to use similar techniques to determine redshifts of galaxies in the Hubble Deep Field (HDF).

In the 1980s, Koo (1985) used photographic plates and 4 filters to observe galaxies, with the help of a colour-colour diagram and theoretical galaxy spectra by Bruzual (1983). He obtained photo-z's for about 100 galaxies, which are correct within an average error of $\Delta z = 0.04$ for $z < 0.35$, where $\Delta z = |z_{\text{spec}} - z_{\text{phot}}|$. This was done by turning the Bruzual templates into colours and plotting lines of constant redshift and varying spectral types. These lines are known as *iso-redshift lines*, as depicted in Figure 8. Other methods which utilised similar colour-redshift relations include the work by Pelló et al. (1996), where they discovered a cluster of galaxies at redshift $z > 0.75$ by looking for an excess in the redshift distribution in the field of a gravitationally-lensed quasar.

3.2.1 *Template Fitting Methods*

The efforts to determine redshifts photometrically were not very motivated during the 1970s due to the advent of spectroscopic surveys, where spec-z's began to be available and were sufficient for research use. However as computing power and optical precision improved, and the emergence of large sky surveys, photo-z researches advanced rapidly. In fact the term 'photometric redshift' only first appeared in Puschell et al. (1982) when he attempted to estimate redshifts of faint radio galaxies using the JHK infrared bands. Puschell was the first to introduce the χ^2 SED template fitting method, it was later improved by Loh and Spillar (1986) which become the template fitting method we know today. With photometry from 6 non-standard passbands recorded

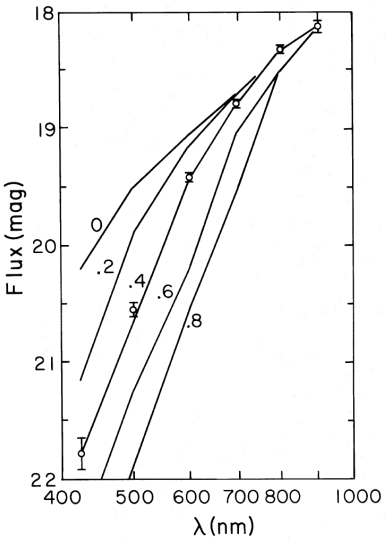

Figure 9: Original figure from Loh and Spillar (1986), which shows the fluxes of an elliptical galaxy through six non-standard filters.

using CCDs, reaching faint limits as deep as $I \sim 21.5$ and redshift $z < 0.5$, they measured photo-z's for 34 galaxies, achieving an average error of $\Delta z = 0.12$.

The concept of the template fitting method is to use a predefined set of typical theoretical galaxy spectra (templates) to fit the photometry of the galaxy, and the photo-z of the galaxy is obtained using the best matched template. The template flux $F_{b,t}(z)$ for template t through filter b at redshift z is obtained by redshifting the SED then integrating it according to its filter response in Equation 39 at rest-frame. Summing through the number of filters n_b used (e.g. 5 *ugriz* bands in SDSS), the observed flux $F_{b,\mathrm{obs}}$ through a filter b for an object is then compared to multiple template fluxes to find a best fit redshift for the object, which can be done by a simple χ^2 test:

$$\chi^2_t(z) = \sum_b^{n_b} \left(\frac{F_{b,\mathrm{obs}} - kF_{b,t}(z)}{\sigma_{F_{b,\mathrm{obs}}}} \right)^2, \tag{68}$$

where k is a normalisation factor and $\sigma_{F_{b,\mathrm{obs}}}$ is the error of the flux. Therefore the best fit photo-z is obtained when $F_{b,\mathrm{obs}}$ matches the template flux with the lowest χ^2 value, and the best fit template also reflects the suggested spectral type of that galaxy.

The template fitting method can be visualised in Figure 9, and original figure from Loh and Spillar (1986). The figure shows an SED of an elliptical galaxy template of Bruzual (1983) through a non-standard filter. The lines depict how the flux of an elliptical galaxy should vary at different redshifts, while the dots showed the observed flux of that particular galaxy, which lies closely to the line of $z = 0.4$. The galaxy was given a photo-z of $z_{\mathrm{phot}} = 0.398 \pm 0.018$.

In the 1990s, many have used this method to obtain photo-z's in the HDF, including the work of Gwyn and Hartwick (1996) and Sawicki et al. (1997), where they studied the evolution of the luminosity function with respect to redshift using different SED templates. Later work focused on exploring the advantages between empirical templates (based on real galaxies) and theoretical templates (based on simulations). There were not many empirical templates at that time since

they are derived from local galaxies, and they do not necessarily represent a wider range of galaxies in terms of luminosity, dust inclination and morphology. Theoretical templates may avoid such problems, however they are based on the assumptions of stellar population synthesis. Koo (1999) showed that the spectra obtained from theoretical and empirical templates can differ in photo-z as large as $\Delta z \sim 0.05$.

From the 2000s onwards, many different template fitting photo-z codes were developed and made publicly available, among which HYPERZ (Bolzonella et al. 2000) was the first template fitting photo-z code freely available for the astronomy community. Other well known template fitting photo-z codes which are still widely used today by various large sky surveys include Photometric Analysis for Redshift Estimate (LEPHARE, Arnouts et al. 1999), Bayesian Photometric Redshift (BPZ, Benítez 2000), Zurich Extragalactic Bayesian Redshift Analyser (ZEBRA, Feldmann et al. 2006) and Easy and Accurate Redshifts from Yale (EAZY, Brammer et al. 2008). Most of these codes differ from one another by the unique features or improvements they have towards the estimation method, but BPZ stands out among all as it was the first algorithm to suggest the use of Bayesian inferences and priors to improve photo-z's, e.g. by assuming that bright ellipticals are unlikely to have high redshifts, the photo-z's could be constrained to reduce the number of catastrophic outliers.

BPZ is used in this work and its functions will be described in detail in Section 3.3.2 and Section 5.4. We will return to the discussion on how template fitting methods have improved in the recent years in Section 3.3.1. For the interest of the reader, a list of publicly available template-fitting photo-z codes can be found in Table 4 at the end of this chapter.

3.2.2 *Empirical Methods*

In the 1990s, a different approach on photo-z estimation was developed. Connolly et al. (1995) was perhaps the first to move away from template based methods to explore empirical equations relating redshift and the broadband magnitudes. They constructed simple quadratic polynomial functions to produce a redshift-magnitude relation $z(m)$:

$$z(m) = a_0 + \sum_{i=1}^{n_b} a_i m_i + \sum_{i=1}^{n_b} \sum_{j=1}^{n_b} a_{ij} m_i m_j, \tag{69}$$

where n_b is the number of filters used, while the constants a_0, a_i and a_{ij} are found by linear regression. They showed that the average redshift error was reduced to $\Delta z \sim 0.05$ for objects with redshifts up to $z \sim 0.5$.

The development of empirical methods for photo-z's was a breakthrough at that time since it marked the start of a data-driven photo-z estimation. This allowed cosmologists to obtain photo-z's while skipping the tedious processes of getting the correct galaxy SEDs, correct theories on galaxy formation and evolution, which are of course important research goals in their own right. Empirical photo-z methods were quickly embraced by the community: SubbaRao et al. (1996) used this method to measure the luminosity function up to $J = 24$; Brunner et al. (1997) used third order polynomials to obtain photo-z's with redshift errors $\Delta z < 0.02$ for redshifts $z < 0.4$ when calibrated via CCD photometry; while Sowards-Emmerd et al. (2000) and Hsieh et al. (2005) used this method to produce photo-z catalogues for SDSS and the Red-sequence Cluster

Survey (RCS) respectively. Wang et al. (2008a) compared this quadratic regression method with linear and cubic regression and concluded that quadratic regression seemed the best and higher order polynomials do not improve the results significantly.

Another important empirical method developed is known as the colour-redshift relation (CZR) for quasars (Richards et al. 2001). The CZR is formulated by first taking a training sample of quasars with known redshifts and having their colours plotted against redshift. A line is then drawn through the mean colours in small bins of redshift. Since they used 4 SDSS colours *u*-*g*, *g*-*r*, *r*-*i* and *i*-*z*, they obtain four CZR plots, and the photo-*z*'s of quasars can be obtained by a simple χ^2 relation:

$$\chi^2(z) = \sum_i^4 \frac{[C_{i,\mathrm{obs}}(z) - C_{i,\mathrm{CZR}}(z)]^2}{\sigma^2_{C_{i,\mathrm{obs}}} + \sigma^2_{C_{i,\mathrm{CZR}}}}, \tag{70}$$

where the $C_{i,\mathrm{obs}}(z)$ denote the four observed colours, while $C_{i,\mathrm{CZR}}(z)$ is the mean colour of the CZR at colour i at redshift z, with σ denoting their uncertainties. The $\chi^2(z)$ for a particular redshift bin is thus summed over all four colours, and the redshift bin with the lowest χ^2 value yields the best redshift for that quasar. This method was later improved by Weinstein et al. (2004), and they showed that 90% of the quasars with photo-*z*'s $0.8 < z_{\mathrm{phot}} < 2.2$ achieve an accuracy of $\Delta z < 0.3$.

3.2.3 *Machine Learning Methods*

With the advancement of computer technology in the late 1990s, empirical methods have expanded to the use of machine learning to estimate photo-*z*'s. Firth et al. (2003) were probably the pioneers in machine learning photo-*z* methods. They used artificial neural networks (ANNs) to obtain photo-*z*'s for about 20 000 galaxies from SDSS, achieving a root-mean-square error $\sigma_{\mathrm{RMS}} \sim 0.021$ for galaxies with redshift $z < 0.35$. Their work later led to the development of the first publicly available machine learning photo-*z* code, Artificial Neural Network Redshifts (ANNZ, Collister and Lahav 2004), which was first used to produce the MegaZ-LRG Photo-*z* Catalogue (Collister et al. 2007), and subsequently many others.

There are many different kinds of machine learning algorithms, but the concept of machine learning shares a similar theme. A machine learning photo-*z* algorithm first needs a *training set*, a set of galaxies with known redshifts and fluxes (or magnitudes) for the multiple filters used. The machine learning algorithm is setup to learn from the training set, creating a set of highly non-linear equations $z_{\mathrm{spec}} = f(m_b)$ to find a relation between the magnitudes m_b for every band b and their respective spectroscopic redshifts z_{spec} for the training set. The machine learning algorithm is treated as a black box, such that the intrinsic equations used are stored internally and need not be known by the user. The training algorithm would constantly update the fitting parameters in these equations, and the training terminates when it reaches a minimum for the *cost function*, which is usually in the form of a difference between the output and the truth value provided by the training set.

A common problem encountered in the training process is *overfitting* (or overtraining), a condition where the training results resemble the training set too closely, so that they fail to predict photo-*z*'s on a different sample reliably. This can be understood graphically through Figure 10,

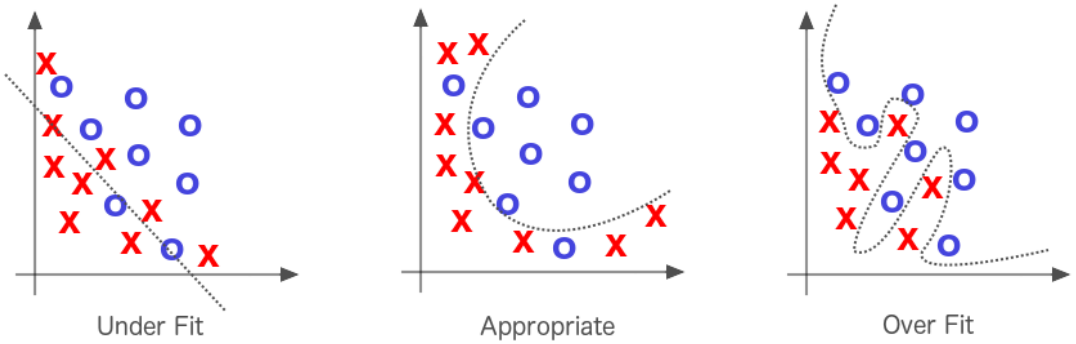

Figure 10: An illustration of overtraining. The dotted line shows a smooth and desired curve needed to separate the blue rings and red crosses (or in our case, different photo-z's) based on machine learning. The left panel shows that the machine has undertrained, the middle panel trained appropriately, while the right panel shows the case when the machine has overtrained: it fitted the data too closely, which may result in a fail when tested on a different sample. Figure obtained from Patterson and Gibson (2017).

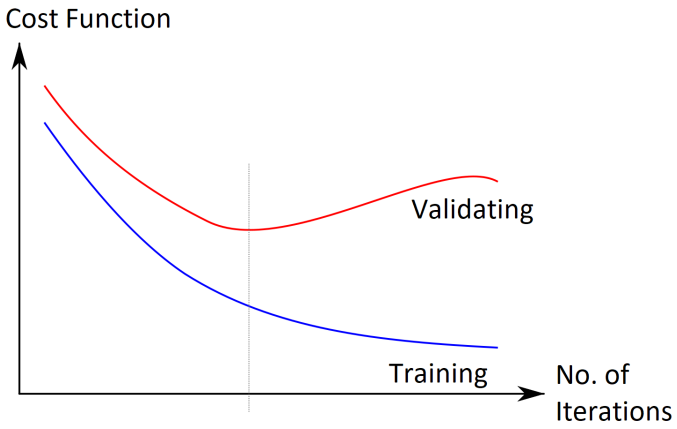

Figure 11: The prevention of overtraining by using a validation set. The training process terminates when the cost function of the validation set reaches a minimum value (dotted vertical line).

where an overfitting gives rise to an overly complicated line to distinguish the crosses and the circles. A simple solution is to use a *validation set*, it is a set different from the training set yet contains known redshift and similar photometric properties to it. This set is trained alongside and constantly compared to the training set, and training terminates when the cost function of the validation set reaches a minimum, as visualised in Figure 11. Other methods to prevent overtraining include the use of cross-validation (using subsets of the training set as validation) and regularisation (see Section 5.2.1).

Once the machine learns the relationship between the magnitudes and the true redshift, it could be run on a *testing set*, which is a different set of objects also with known redshift and similar photometric properties to the training and validation set. Photo-z's are produced on this set, they are then compared to their spec-z's using different metrics to evaluate the performance of this estimation. The machine learning algorithms used usually contain several *hyperparameters* or random seeds which could be tweaked or run multiple times and averaged to produce better results. Once the performance is satisfactory, the machine learning algorithm is run on the *target*

set, which is a set of galaxies with no redshift information available, and the photo-z's obtained from such a set are compiled to produce a photo-z catalogue for a survey, ready to be used for weak lensing or large-scale structure analysis.

From a Bayesian statistical point of view, the priors on the underlying redshift distribution are automatically taken into account with the use of training data, however this is only true if the training data is representative of the sample where photo-z's are calculated on. Machine learning codes can in fact be trained on simulated data (Vanzella et al. 2004; Wadadekar 2005) or samples with data augmentation (Hoyle et al. 2015b), both have shown to achieve better results. Machine learning algorithms also benefit from the fact that we could easily add other information like galaxy size or concentration to the inputs as long as the computing power permits. Therefore machine learning algorithms quickly became powerful tools to obtain photo-z estimates for large sky surveys when large training sets are available.

Ever since the first machine learning photo-z algorithm ANNZ debuted in 2004, many photo-z codes based on different machine learning algorithms have been developed. The number of publicly available machine learning photo-z codes have reached 20 in the past 15 years. Popular machine learning algorithms other than the ANN used by the photo-z community include support vector machines (SVMs), Gaussian processes (GPs), K-nearest neighbours (KNNs) and boosted decision trees (BDTs). All these machine learning algorithms differ from one another in many ways, for instance, generally ANNs fare better with more inputs used, while SVMs require an optimal number of inputs for best performance (Wang et al. 2008b); the performances of BDTs and GPs are highly dependent on the hyperparameters set, while SVMs do not require any architecture (Hoecker et al. 2007).

There are too many machine learning algorithms and codes available on the market to be discussed individually in this work. We leave with the reader Table 3, a list of machine learning algorithms arranged in chronological order according to the year they were first used for photo-z estimation, together with the papers describing them. However, the algorithms employed in this work (ANN, BDT, KNN and GP) will be further discussed in Chapter 5. The reader will also find a list of publicly available machine learning photo-z in Table 4 at the end of this chapter, of which the codes ANNZ (Collister and Lahav 2004), ANNZ2 (Sadeh et al. 2016) and DELIGHT (Leistedt and Hogg 2017) have been used in this work, and will also be described in Chapter 5.

3.2.4 *Template Fitting v.s. Machine Learning*

Both template fitting and machine learning photo-z codes are widely used in many current and upcoming galaxy surveys. Machine learning photo-z methods often produce better photo-z overall when compared to template based methods (Hildebrandt et al. 2010). Simple photo-z v.s. spec-z plots of machine learning codes typically show lower scatter and outlier fraction than template codes. However machine learning codes tend to produce a lower density of objects close to the diagonal line in the spec-z v.s. photo-z plot, or in other words, they produce larger 68th percentile errors (σ_{68}, see Section 5.7.1 for definition). Machine learning codes also produce better photo-z's at intermediate redshifts where training objects are plentiful, while template fitting codes fare better at the lower or higher end of the redshift limit (Abdalla et al. 2011).

Table 3: List of machine learning algorithms arranged chronologically according to the year they were first introduced to estimate photo-z's. Since each machine learning technique spur the development of many photo-z codes, the source only lists the first paper it debuted in the photo-z community.

Machine Learning Algorithm	**Introduced By**
Artificial neural networks	Firth et al. (2003)
Support vector machines	Wadadekar (2005)
Gaussian process regression	Way and Srivastava (2006)
K-nearest neighbours	Ball et al. (2007)
Kernel regression	Wang et al. (2007)
Spectral connectivity analysis	Freeman et al. (2009)
Random forests	Carliles et al. (2010)
Boosted decision trees	Gerdes et al. (2010)
Weak gated experts	Laurino et al. (2011)
Self organising maps	Way and Klose (2012)
Relevance vector machine	Sánchez et al. (2014)
Symbolic regression	Krone-Martins et al. (2014)
Generalised linear models	Elliott et al. (2015)
Genetic algorithms	Hogan et al. (2015)
Ordinal class PDF estimation	Rau et al. (2015)
Deep neural networks	Hoyle (2016)
Directional neighbourhood fitting	De Vicente et al. (2016)
Sparse Gaussian process framework	Almosallam et al. (2016a)

Both template and machine learning photo-z codes have their disadvantages. Template photo-z estimation is primarily limited by the templates used: we need to assume that the templates used are complete and representative of the sample of galaxies that we are interested in, and that the filter responses are well calibrated. Massarotti et al. (2001) showed that for low redshifts, the choice of templates have minor effects on the performance as long as a starburst galaxy templates are present; however for high redshifts, its performance relies heavily on how well physical processes like photon absorption in the interstellar and intergalactic medium are implemented. Template based photo-z methods are also not as flexible as machine learning methods when it comes to adding extra information, since machine learning codes could easily allow extra input parameters like galaxy size, axis-ratio and Sérsic index be added in training to further improve the photo-z's. Besides, systematic side-effects associated with the photometry (e.g. zero-point errors of magnitudes) have to be dealt with manually and could be difficult to handle.

On the other hand, machine learning methods suffer from the fact that they require a complete and representative training set to work: they do not perform well when extrapolated. Since most of the training galaxies have brighter magnitudes and lower redshifts, they are not representative to train for samples of galaxies with high redshifts and faint magnitudes (a workaround to this problem is discussed in Section 3.3.4). Besides, when the number of inputs are significantly increased (e.g. multiple narrow band filters / many morphological parameters), machine learning algorithms tend not to produce a tight fit when compared to template codes, which we will see later in Chapter 8 as part of the results of this work.

Therefore template-based and machine learning-based methods are still used side-by-side and compared today, since the results from either method can be selected based on the need of the research. For a general rule of thumb, template-fitting methods are preferred when exploring new regimes, while empirical methods are preferred when large training sets are available and when high precision is required (Zheng and Zhang 2012). Efforts to combine photo-z's obtained from both methods have been explored (Dahlen et al. 2013; Cavuoti et al. 2017a), and codes to incorporate both methods have been produced (Wolf 2009; Leistedt and Hogg 2017). In fact Budavári (2009) created a unified framework for photo-z's where a rigorous mathematical solution was produced. Through this he pointed out that the two methods are in fact special cases of such a generalised framework which, if well studied, would provide a clear direction of improvement on whichever code one is working on. Efforts on combining and comparing photo-z results from different codes will be further discussed in Section 3.3.5.

3.3 Efforts to improve photometric redshifts

Thus far we have seen the importance of photo-z research, the different methods, and their advantages. As photo-z is still an active field of research, it is worth studying recent efforts to improve photo-z's, from code to result. In the following sections we summarise 7 important areas in which photo-z estimations have been improved and developed on.

3.3.1 *Improving Galaxy SED Templates*

Since template photo-z's are highly dependent on the templates used, it is natural to work on improving the galaxy SED templates used for estimation. We mentioned earlier how BPZ (Benítez 2000) used Bayesian priors to reduce the number of outlier photo-z's. They also assert that the redshift precision can be improved by using interpolation between different templates. In fact this was done by LEPHARE, where they produced new templates by optimising the original templates of Coleman, Wu and Weedman (or the CWW template set, Coleman et al. 1980) and one starburst template by Kinney et al. (1996) using a spectroscopic set of galaxies from the VIMOS VLT Deep Survey (VVDS, Le Fèvre et al. 2005) to produce a linearly interpolated 62 templates. These templates are commonly known among the photo-z community as the LEPHARE template set (more on this in Section 5.6.3), combined with zero-point offset refinements and adding a Bayesian prior, they have shown that catastrophic outliers have been reduced by a factor of 2 (Ilbert et al. 2006). The same group of people did a similar job for the COSMOS photo-z catalogue (Ilbert et al. 2009), but this time using 9 templates from Polletta et al. (2007) and 12 templates from Bruzual and Charlot (2003), and these templates are known as the COSMOS template set (Section 5.6.4).

Babbedge et al. (2004) were the first to explore the improvement of photo-z's by using a combined template set of galaxies and quasars. They tested their code IMPZ on galaxies from the Isaac Newton Telescope Wide-Angle Survey (INT WAS, McMahon et al. 2001) and found improvement in the results. Assef et al. (2008) constructed low resolution spectral templates using galaxies and AGNs, and estimated photo-z's with their code LRT on a sample of galaxies from

the AGN and Galaxy Evolution Survey (AGES, Kochanek et al. 2012). Their results achieved a photo-z error of $\Delta z/(1+z) < 0.04$.

Other notable improvements in template photo-z algorithms include GAZELLE (Kotulla and Fritze 2009) which derives its templates from chemically consistent galaxy evolutionary synthesis models; Great Observatories Origins Deep Redshift (GOODZ, Dahlen et al. 2010) uses spec-z's to correct offsets between template and observed SEDs and saw improved results; and SED Analysis Through Markov Chains (SATMC, Johnson et al. 2013) produces novel SED fitting routines using the Markov Chain Monte Carlo (MCMC) technique. The more recent and important addition to the template community was the production of the Brown templates (Brown et al. 2014), which are made based on real galaxies and have been shown to produce excellent photo-z results (see Section 5.6.5 for more details).

3.3.2 *Photometric Redshift PDFs and Distributions*

The production and use of photo-z probability density functions (PDFs or $p(z)$) has become almost a must in recent years. It is intuitive that the full $p(z)$ would contain more information than a mere point estimate alone. For instance, the $p(z)$ may be bimodal with its secondary peak value closer to the true redshift, which collectively would yield a more accurate redshift distribution $n(z)$ for the sample. Therefore this section summarises the history, development and importance of photo-z PDFs.

Photo-z PDFs were first introduced by SubbaRao et al. (1996), where they set every galaxy to have a Gaussian PDF with mean redshift z and standard deviation (or photo-z error) σ. In order to produce a continuous estimate of the luminosity function, the photo-z PDFs of each galaxy were translated into PDFs in absolute magnitude. This was a breakthrough at that time since galaxy samples were small and did not have smooth redshift distributions as we have today.

The success of photo-z PDFs led codes like BPZ (Benítez 2000) and LEPHARE (Arnouts et al. 1999) to incorporate photo-z PDFs as part of their output products. BPZ constructed its PDF differently from SubbaRao et al. (1996), and it could be understood with the help of Figure 12. First we have the likelihood function for each template $p(m|z,t)$, the probability a galaxy would have magnitude m given redshift z for different templates t (panel 1). Then we have the prior $p(z,t)$, the probability distribution of z for different templates t (panel 2). The prior and the likelihood are multiplied to get the posterior distribution $p(z,t|m)$ of redshift z for each template t (panel 3), in which we can see that the curve on the right is heavily suppressed since the probability that a galaxy has such a high redshift for that template is very low. The final posterior distribution $p(z|m)$ is produced when summed (i.e. marginalised) over all templates (panel 4).

BPZ also introduced a *Bayesian odds parameter* (denoted as Θ throughout this work) to characterise the 'peaky-ness' of the PDF. Θ is defined as

$$\Theta = \int_{z_{\rm peak}-\delta z}^{z_{\rm peak}+\delta z} p\left(z|m_b\right) dz, \tag{71}$$

where $p(z|m_b)$ is the PDF distribution of the output photo-z, m_b refers to the list of magnitudes for a number of bands b, and z is the true redshift. The integral measures the probability mass between the values $z_{\rm peak} \pm \delta z$, where $z_{\rm peak}$ is the peak of the PDF, and $\delta z = k(1+z_{\rm peak})$. Θ

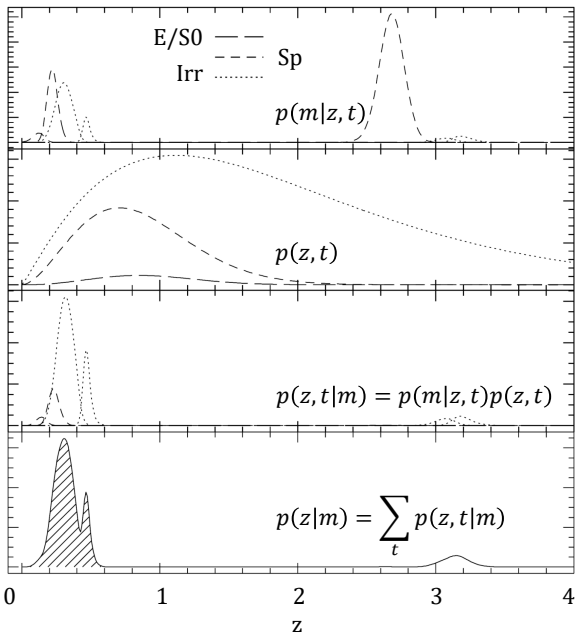

Figure 12: From top to bottom: the likelihood function $p(m|z,t)$ for each template, the prior probability $p(z,t)$ for each template, the probability distribution $p(z,t|m)$ for each template, and the final Bayesian probability $p(z|m)$ summed over each template. Figure obtained and modified from Benítez (2000).

ranges between 0 and 1, with values closer to 1 implying that the photo-z obtained is more reliable, therefore the constant k is arbitrarily chosen such that not too many objects end up having $\Theta = 1$. Benítez (2000) set $k = 3 \times 0.067$, while in this work we used the same value when using BPZ but setting $k = 1 \times 0.067$ when using ANNZ2 (more on this in Chapter 5). Θ is used as a quality cut in many photo-z catalogues, among which Margoniner and Wittman (2008) mentioned that a cut in Θ is better than a cut in S/N since the latter would cut away a lot of usable and faint galaxies which are crucially needed for weak lensing analyses.

The PDF modelled by Benítez (2000) was later improved by Fernández-Soto et al. (2002) where they used a likelihood function which encapsulates the photometric and systematic errors. PDFs produced by machine learning codes on the other hand are a little different as they do not have priors and templates to begin with. Ball et al. (2008) derived photo-z PDFs by using a KNN algorithm, whereby the algorithm compares the position of that galaxy in magnitude space with respect to K number of neighbours in the spectroscopic training set and derives the error of the photo-z, which translates into the PDF.

Sadeh et al. (2016) uses the same method for their code ANNZ2, however the PDFs are averaged over different runs using different machine learning methods to obtain better estimates. Polsterer et al. (2016) on the other hand introduced a PDF generation method using random forests, where the magnitude space is partitioned by independent decision trees (see Section 5.3.1), and the PDFs are constructed by comparing with objects from the spectroscopic sample within the partition assigned.

Many have advocated the use of $p(z)$ to replace the use of point estimates to represent photo-z's of individual galaxies and quasars. Ball et al. (2008) showed that they virtually eliminated all catastrophic quasar photo-z's observed in the optical for samples from SDSS and the Galaxy

Evolution Explorer (GALEX, Martin et al. 2005) by post-selecting $p(z)$ with a single peak, with Wolf (2009) agreeing to this result. Instead of using the peak of the $p(z)$ as the best point estimate, Wittman (2009) used random numbers between 0 and 1 to select a Monte Carlo sample of the normalized cumulative $p(z)$ of a galaxy, with the corresponding redshift $z_{\rm MC}$ selected as the best point estimate. This was done for a large sample of galaxies, and he showed that a plot of $z_{\rm MC}$ v.s. $z_{\rm spec}$ showed less scatter than $z_{\rm peak}$ v.s. $z_{\rm spec}$, further suggesting that $p(z)$ would reduce the systematics in weak lensing analysis. Tanaka et al. (2018) also introduced a new point estimator. They defined a 'risk parameter' as a function of redshift and chose the point where the risk is minimised as the best point estimate. This was done by integrating the $p(z)$ with the loss function $L(z)$, an 'inverted Lorentz function' which can be viewed as a loss arising from $z_{\rm phot}$ being different from $z_{\rm spec}$. They demonstrated that this point estimate resembles $z_{\rm spec}$ more than the mean, mode or median of the $p(z)$.

When the $p(z)$ of each galaxy in a sample are stacked together, it produces a smooth $n(z)$ distribution, and it has been shown to be a better $n(z)$ representation than using histograms of the photo-z point estimates. Mandelbaum et al. (2008) have tested the use of full $p(z)$ for weak lensing studies, and find that it tends to give superior results than using the point estimates alone, with calibration biases close to zero for all lens redshift distributions. The work of Gerdes et al. (2010) and Bonnett (2015) also agreed with this results, showing that the stacked PDFs yield a more accurate reconstruction of both the overall and binned redshift distributions. In cases when the produced photo-z PDFs are too narrow, Dahlen et al. (2013) suggested a simple way to broaden them by smoothing each point in the PDF with a combination of adjacent bins through a simple equation $p(z_i) = 0.25p(z_{i-1}) + 0.5p(z_i) + 0.25p(z_{i+1})$, and they showed improved results.

The use of $p(z)$ however comes with a storage problem. For example, outputting a linearly binned $p(z)$ with a binning of $\delta z = 0.01$ for $0 < z < 2$ would require 200 times more storage space than a single point estimate, and the number increases if a higher resolution PDF is required. Carrasco Kind and Brunner (2014b) attempted to solve this storage problem by using an orthogonal matching pursuit algorithm, such that each $p(z)$ can be represented by a combination of Gaussian and Voigt basis functions. In this way, the number of points needed for storing a PDF at a binning of $\delta z = 0.01$ can be reduced by at least a factor of 10 to 20. Malz et al. (2018) presented a code Qp and used it to investigate the performance of 3 photo-z PDF compression formats (regular quantiles, random sampling and step functions). They conclude by recommending the use of regular binning in probability rather than redshift when storing photo-z PDFs for future surveys.

The increased use of $p(z)$ by the photo-z community also invited the question of how to characterise or evaluate the performance of photo-z PDFs, since there is no such thing as a 'true' redshift PDF to compare with. Polsterer et al. (2016) introduced two metrics: the continuous ranked probability score (CRPS) and the probability integral transform (PIT), both widely used by the weather forecast community. The CRPS is a score that has the same unit as the variable of interest, the lower the score the better the $p(z)$ captures $z_{\rm spec}$ within its curve. The shape of the PIT histogram on the other hand, would tell if the overall $p(z)$ of galaxies are too broad or narrow. Both PIT and CRPS will be used in this work and further described in Section 5.7.2.

Wittman et al. (2016) introduced the highest probability density credible interval to study overconfidence of PDFs, testing if the PDFs produced have the 'correct width'. This method uses the principle that a sample of galaxies should contain 1% of galaxies whose $z_{\rm spec}$ is in tension

with z_{phot} at the 99% level, and the overall $p(z)$ is considered overconfident if too many galaxies exhibit such tension. This test was conducted on $p(z)$'s produced by BPZ and EAZY and they concluded that the PDFs produced by BPZ are generally overconfident. In an upcoming paper, Schmidt et al. (in prep.) will compare the performances of overall $p(z)$ and $n(z)$ of 11 codes using several metrics in preparation for LSST, to which part of the work in Section 7.4 contributed to.

3.3.3 *Improving Photometric Redshift Errors*

Photo-z errors play an important role in telling us the reliability and accuracy of a galaxy's photo-z value in a catalogue. At best they should be an indicator to remove low quality redshifts from a sample, thus they should be formulated in such a way to correlate with $|z_{\text{spec}} - z_{\text{phot}}|$. Fernández-Soto et al. (2002) suggested that photo-z errors should be obtained through their respective photo-z PDFs, the Bayesian odds from BPZ as mentioned earlier is one example. However Hildebrandt et al. (2008) ran BPZ and HYPERZ on multiple imaging datasets with different depths and filter coverages, and they concluded that photo-z error estimates from these codes do not correlate tightly with the accuracy of the photo-z's. Dahlen et al. (2013) compared the performance of 9 template codes and found that most photo-z codes underestimate their photo-z errors. Since in most cases the uncertainty of magnitudes only represent the Poissonian noise on the corresponding photon-count, they do not take into account systematic uncertainties and correlations between observables and are not always precise in itself. This is bad news since these photo-z errors would propagate and affect the results of the user, most prominently in weak lensing studies as highlighted by Huterer et al. (2006).

A prominent work to improve error estimation was done by Oyaizu et al. (2008b). First they divided photo-z error estimation into two categories: statistical errors from template fitting (χ^2 errors) and propagated errors from input parameters (magnitude errors). They showed that training-set based errors are better quality cuts compared to template errors when a representative training set is available. They also introduced two new error estimators: one using kd-trees to bin data in magnitude space and finding $|z_{\text{spec}} - z_{\text{phot}}|$ at the median of that point, and the other using nearest neighbours (as mentioned in Section 3.3.2, and further explained in Section 5.3.2), where the latter performed better than kd-trees and the propagated errors.

Carliles et al. (2010) studied the error distribution of photo-z's. They showed that using random forests to estimate photo-z's produces reliable error distribution estimates even on data with highly skewed noise distributions. However, the nearest neighbour method for error determination is currently the best choice, widely used by many codes and current surveys.

3.3.4 *Representativeness of Training Sets*

Spectroscopic galaxy samples are usually constructed by cross-matching photometry from large photometric surveys with redshifts obtained from multiple different spectroscopic surveys. Due to the different target selections of each spectroscopic survey, the combined spectroscopic training sample will contain objects which are unevenly distributed in colour-magnitude space, and are preferentially brighter and redder galaxies. This means that the distribution of training parameters (e.g. magnitude, colour, size etc.) in the training, validation and testing samples will

turn out to be quite different from that of the target sample, in which the photo-z's are to be estimated. This is a problem for machine learning methods since they generally do not perform well when extrapolated, and will induce biases in template codes if they use non-representative validation samples for calibration.

Lima et al. (2008) attempted to solve this problem by reweighting the training set. Reweighting is done by comparing the density of objects in a selected parameter space (usually the colour and magnitude) of the training and target samples, and setting a weight value to each object in the training sample. The weights are estimated using a nearest neighbour technique that ensures stability in sparsely populated regions of colour-magnitude space. With the weights in hand, the redshift distribution $n(z)$ of the target sample can be estimated by weighting the redshift distribution of the training set. The key assumption behind this method is that two samples with identical distributions magnitude and colour should have identical $n(z)$.

Initially the work of Lima et al. (2008) was meant to produce an estimated redshift distribution for the target set without deriving photo-z's. Later a follow up paper by Cunha et al. (2009) showed that the $p(z)$ for individual galaxies in the target sample can be derived by summing the weighted redshift distribution with its nearest neighbours in the training set. Sánchez et al. (2014) later used this weighting method to evaluate the targeted photo-z performance on a set that mimics the full DES photometric sample. This is done by using weights in the training process, such that the cost function used to estimate the photo-z will be balanced by upweighting objects that are less represented in the training sample compared to the target sample, and downweighting objects otherwise. The final results on the testing set are also reweighted to reflect the expected photo-z performance on the deeper full sample. This technique is also used in this work when evaluating the impact of morphology on photo-z's on the CS82 sample in Section 6.5.1.

3.3.5 *Comparing Different Photometric Redshift Algorithms*

With the huge number of publicly available photo-z codes and the large variety of estimation algorithms, at the very least one might wish to know which code performs the best. Of course the different codes do not always appear to be in competition with one another, in fact most of the time they act as checks or verifications that a code is performing relatively well. The first collective effort to compare the performance of different photo-z codes was conducted by Hogg et al. (1998). It was a blind test conducted to end scepticism among the astronomy community towards the reliability of photo-z's at that time. The test set was a sample of 27 galaxies from the HDF with redshift $z < 1.4$, with 5 universities (Victoria, Imperial College, Toronto, John Hopkins and Stony Brooks) participating in the test using a variety of template fitting and empirical methods. All codes resulted in a redshift accuracy of $\Delta z < 0.1$ for more than 68% of the sources, which led to the consensus that the photo-z method is indeed reliable to use.

Later code comparisons focused on comparing the performances of different machine learning algorithms. Wang et al. (2008b) compared the performances of SVM and kernel regression with the latter performing better, despite SVM being faster to run. Bonfield et al. (2010) compared the performances between GP and ANN and showed that GP is successful at interpolating through an artificially created redshift gap of redshift $1.3 < z < 1.7$ (mimicking the redshift desert), while

ANN suffered strong bias. In the process they also found that inaccurate redshifts contaminating the training sample would have little effect on both codes, and that a minimum training size of 2000 objects is suggested for satisfactory results.

Later Hildebrandt et al. (2010) initiated the Photo-z Accuracy Testing (PHAT) programme, an international initiative to test and compare different photo-z codes on a fixed sample of simulated and real galaxy data. 17 codes including BPZ, EAZY, GAZELLE, GOODZ, HYPERZ, KERNELZ (Wolf 2009), LEPHARE, LRT, ZEBRA, ANNZ and many others participated in the test, and their performances were ranked by scatter, bias and outlier rate fraction. Good agreement was seen among each code when tested on a highly idealised simulation, however it was found that most template codes do not fare well when tested on real data from the Great Observatories Origins Deep Survey (GOODS, Giavalisco et al. 2004) when infrared photometry was included, suggesting considerable inaccuracy in the SED templates used in the infrared region. With adequate catalogue cleaning, an outlier rate below 5% was achieved for some codes.

As mentioned previously, the general conclusion when comparing different codes is that the performance of each code depends on the figure of merit used and that no single best photo-z code exists. This statement is confirmed by the work of Abdalla et al. (2011) when comparing 8 photo-z codes to produce the updated MegaZ-LRG Photo-z Catalogue, and the work of Dahlen et al. (2013) when comparing 9 template fitting codes using different SEDs and priors on data from the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS, Grogin et al. 2011).

Instead of searching for the best photo-z code, many have turned their focus to enhancing photo-z's by combining the strengths of different codes. Dahlen et al. (2013) explored 3 different methods to combine photo-z results: taking the median photo-z, averaging the $p(z)$, and using a hierarchical Bayesian method from each of the 9 codes, and found that the photo-z's produced are very similar but significantly decrease the scatter and outlier fraction when compared to the photo-z's produced individually. Cavuoti et al. (2017a) used 5 different codes including Multilayer perceptron quasi Newton algorithm (MLPQNA, Brescia et al. 2014), LEPHARE and BPZ to obtain photo-z's for KiDS, they combined the photo-z of each code by first separating them by their spectral type (E, S0, Sab, Scd and SB) using a template code, and the final photo-z is obtained by the training code MLPQNA, thus combining the best of both worlds. The final results showed a σ_{68} reduced from 0.018 to 0.016, however with its bias increased. Duncan et al. (2017) and Duncan et al. (2018) combined template and GP photo-z's via hierarchical Bayesian combination, and the hybrid photo-z is shown to reduce the catastrophic outlier fraction by a factor of 4.

Sánchez et al. (2014) compared the performances of 13 codes including DESDM (Oyaizu et al. 2008a), ANNZ, Trees for Photo-z (TPZ, Carrasco Kind and Brunner 2013), ANNZ2, ARBORZ (Gerdes et al. 2010), SKYNET (Graff et al. 2014), BPZ, EAZY, LEPHARE, ZEBRA and PHOTOZ (Bender et al. 2001) on the Science Verification Data of DES, incorporating reweighting to study a forecast on their performances on the full DES photometric sample. They concluded that while all codes comfortably meet the photo-z requirements of DES, empirical codes generally perform better, achieving a scatter of $\sigma_{68} \sim 0.08$. Tests on a deeper subsample showed that the impact of the u-band in this case is not as great as in previous shallower surveys like the SDSS. The follow up work by Bonnett et al. (2016) used 4 codes (ANNZ2, TPZ, SKYNET and BPZ) to present galaxy photo-z estimates for weak lensing analyses in the DES Science Verification Data.

Perhaps the latest major work on comparing codes was the photo-z catalogue produced by Tanaka et al. (2018) for the HSC Subaru Strategic Program DR1. They used 6 codes including Direct Empirical Photometry (DEMP, Hsieh and Yee 2014), Flexible Regression over Associated Neighbours with Kernel Density Estimation for Redshifts (FRANKEN-Z, Speagle et al., in prep.), Extended Photometric Redshift (EPHOR), Self Organizing Maps Redshift (SOMZ, Carrasco Kind and Brunner 2014a) and MIZUKI (Tanaka 2015) to obtain photo-z's for the programme, and they find all codes achieve $\sigma_{\rm RMS} \sim 0.04$ and 8% outlier fraction rate for the redshift range $0.2 < z < 1.5$ and magnitude $i < 23$. An upcoming photo-z comparison (Schmidt et al., in prep.) will assess photo-z PDFs in the context of LSST.

3.3.6 *Adding Extra Information*

Another way improvement in photo-z's can be achieved is by adding extra information on top of magnitudes and colours to the training process. The training of different types of magnitude fits have been done (Wang et al. 2010) but the improvement is small or almost negligible. It is expected that extra physical information of galaxies, usually morphological properties like size, ellipticity or concentration would add an extra dimension to the training process and break degeneracies of photo-z's in certain redshift regions. In one way, it would help us to discover redshift dependencies on galaxy morphologies too.

For template fitting methods, many have included the use of surface brightness as a prior in SED templates (Kurtz et al. 2007; Stabenau et al. 2008; Xia et al. 2009), motivated by the knowledge of surface brightness dimming at a rate of $(1+z)^4$. In empirical methods this application is more straightforward, since algorithms are constructed such that it is not difficult to add extra input parameters to the training. For example, Collister and Lahav (2004) and Wadadekar (2005) demonstrated that by including the 50% and 90% Petrosian flux radii ($r_{\rm P50}$ and $r_{\rm P90}$), the photo-z root-mean-square error $\sigma_{\rm RMS}$ improved by 3% and 15% respectively for the SDSS MGS. Tagliaferri et al. (2003) used Petrosian fluxes and radii in their work on galaxies from the SDSS Early Data Release and calculated robust errors decreasing as much as 24%. Meanwhile, Vince and Csabai (2006) included the concentration of galaxy light profiles in their study and reported that the $\sigma_{\rm RMS}$ on SDSS galaxies improved by 3%. Wray and Gunn (2008) included surface brightness and the Sérsic index, and found improvements in variance when compared to other template fitting methods applied to the SDSS MGS previously.

A particularly thorough investigation was performed by Way et al. (2009), who studied how galaxy morphology information affects photometric redshift quality. Using GPs, they included several galaxy morphological parameters alongside with photometry for training, like $r_{\rm P50}$, $r_{\rm P90}$, concentration index C, `fracDeV` f and the Stokes Q and U parameters (see Section 2.2.2 for definitions). They showed that the addition of these parameters does not systematically improve the photo-z estimation. Later, Way (2011) separated galaxy samples into ellipticals and spirals with the help of Galaxy Zoo (Lintott et al. 2011), and showed that photo-z estimated using adaptive moments and texture for the SDSS LRGs yield $\sigma_{\rm RMS} \sim 0.012$.

Singal et al. (2011) formed principal components of a series of 8 derived morphological shape parameters (smoothness, asymmetry, Gini coefficient etc.) and used these in combination with photometry to improve photo-z estimations. However, they found that outliers were not signifi-

cantly decreased, and the shape parameters may have contributed noise instead. Jones and Singal (2017) repeated the study with their SVM code SPIDERZ, and obtained results in agreement with the earlier work.

In recent literature, Gomes et al. (2018) used GPZ, an algorithm using GP to explore methods of improvement in photo-z's. Working on photometry from SDSS and United Kingdom Infrared Deep Sky Survey (UKIDSS, Lawrence et al. 2007), and spec-z's from Galaxy and Mass Assembly (GAMA, Liske et al. 2015), they included near-infrared photometry and angular size as inputs and saw an improvement of up to 20% in photo-z accuracy.

3.3.7 *Other Improvement Methods*

On top of the photo-z improvement efforts mentioned above, the following methods do not fall into the above categories:

- *Upgrading codes.* ANNZ has been succeeded by ANNZ2, which now supports the reweighting of training sets, PDFs and the inclusion of other machine learning methods. PHOTORAPTOR (Cavuoti et al. 2015) became the successor of MLPQNA, it has a user interface built upon the java environment. The same team later made Machine-learning Estimation Tool for Accurate Photometric Redshifts (METAPHOR, Cavuoti et al. 2017b), which uses ANNs, Random Forest and KNNs to produce reliable photo-z PDFs and error distributions.

- *Metric representation.* Carrasco Kind and Brunner (2014a) introduced the 'I-score', a metric which efficiently takes into account different metric indicators like the scatter, bias, outlier and the Kolmogorov-Smirnov (KS) statistic (see Section 5.7.3) by normalising each metric and compiling a weighted sum. They find it easier to use one metric for multiple photo-z code comparison.

- *Anomaly removal.* Hoyle et al. (2015a) used the elliptical envelope method to remove anomalies in the galaxy training sample before the photo-z estimation process. The routine models the data as a high dimensional Gaussian distribution and finds a boundary ellipse which contains most of the data. Any data outside of the ellipse are classified as anomalous and removed, the photo-z estimated without the anomalies showed great improvement across all tested metrics, reducing the outlier fraction up to half for some codes tested.

- *Stacking.* Zitlau et al. (2016) introduced stacking, whereby the output photo-z is fed back to the machine learning algorithm together with the magnitudes to reproduce a new photo-z output. They showed that for strong learners (ANNs or BDTs), improvement of the metrics can reach up to 2.5%.

- *Filter dependency.* In preparation for the Chinese Space Station Optical Survey (CSS-OS), Cao et al. (2017) found that the *gri* bands have the largest impact on photo-z scatter; NUV and u bands have great impact on outlier fractions; zy bands have relatively small impact; while JH bands improve photo-z's at redshift $z > 4$. This was done by using LEPHARE on COSMOS galaxies, involving galaxies up to $i \sim 26$.

- *Feature selection.* D'Isanto et al. (2018) studied the possibilities of efficient automatic feature selection to optimise photo-z performance using random forests and deep convoluted net-

works (D'Isanto and Polsterer 2018). Selecting from over 4000 synthetically created features derived from galaxy magnitudes and morphology, they demonstrated that their algorithm improve the photo-z performance compared to using the traditional magnitude inputs.

The research on photo-z is still actively ongoing and is showing no sign of slowing down. Brescia et al. (2018) nicely summarised the 3 current important research questions in photo-z research:

1. The problem of missing and incomplete data,
2. The definition of an optimal input parameter space and feature selections in terms of completeness and accuracy of results, and
3. The derivation of reliable photo-z PDFs.

With the availability of computing tools and advanced telescopes being built, it is indeed an exciting time to do research in photo-z's. We look forward to what new breakthroughs we could achieve with the completion of upcoming large sky surveys.

3.4 MOTIVATION OF THIS RESEARCH

After the discussion on photo-z's in the previous few sections, it is now appropriate to state the motivation of this research work. As expected, this work will focus on improving and enhancing the performance of photo-z's, however since there are many areas of photo-z's that one could work on, this work focuses on 3 main areas, which we will discuss in detail below.

(A) *Photo-z and morphology*. It is natural to expect that photo-z's estimated for different surveys with different redshift and magnitude ranges will benefit from galaxy morphology information to different degrees. However, due to different results reported by different groups as we have seen from Section 3.3.6, a comprehensive study on this subject is warranted. Many current and upcoming surveys (particularly researches that focus on weak lensing) could definitely have their photo-z estimation methods benefit from the high fidelity galaxy morphological parameters that would come for free. Therefore Chapter 6 of this work is aimed at studying the extent to which different morphological parameters improve photometric redshifts, what kind of objects or photometric conditions benefit most from it. This study is different from previous studies in that it is conducted in a more comprehensive manner: using data sets with high quality morphology, varying the number of filters, using photometry of different qualities, testing individual morphological parameters and considering the inclusion of quasar spectra. This chapter will answer the following research questions:

1. Which galaxy morphological parameters improve photo-z's the most in the context of (a) SDSS and (b) the CS82 Survey?
2. What are the different impacts of galaxy morphology on (a) a bright v.s. faint sample, (b) ellipticals v.s. spirals, and (c) LRG-liked samples?
3. What are the impacts of galaxy morphology on photo-z's on galaxy samples under sub-optimal conditions, namely (a) low-quality photometry, (b) fewer than 5 broadband filters, and (c) contaminated by quasars?

(B) *Photo-z PDFs and distributions*. Following the scientific needs to use high quality photo-z PDFs and $n(z)$, Chapter 7 of this work continues the effort of the previous section to explore the impact of galaxy morphology on $p(z)$ and $n(z)$. At the same time, we also explore the use of a few performance metrics to characterize the quality of $p(z)$ and $n(z)$. This chapter is completed as part of a collaborative work with the CS82 Survey (Moraes et al. 2014) and the LSST-DESC PZ Working Group. A photo-z catalogue was produced for CS82 (Soo et al. 2018), and photo-z PDFs using ANNZ2 and DELIGHT have been produced and submitted to the LSST-DESC PZ group to be assessed with other codes, and final results will be featured in an upcoming paper (Schmidt et al., in prep.). As part of the publication policy of LSST, since the work of the latter is not yet finalized and published, only original work of the author will be featured in this section. This chapter will answer the following research questions:

1. What is the impact of galaxy morphology on the $p(z)$ of individual galaxies and on the overall $n(z)$ in the context of the CS82 Survey?

2. How viable are morphological redshifts (morpho-z): redshifts produced using only one photometric broadband and multiple morphological parameters, in the context of the CS82 Survey?

3. How does the performance of ANNZ2 compare to DELIGHT in the context of LSST, and do the photo-z estimates and PDFs produced by both algorithms fulfil the requirements of LSST?

(C) *Photo-z in the context of the narrowband surveys*. Photo-z's face different challenges in narrowband surveys. As the number of bands increase 8 fold, the problem of missing and incomplete data becomes prominent, and the performances of machine learning codes dealing with high dimensional inputs are challenged. Chapter 8 of this work is part of a collaboration with the PAU Survey, we use simulated catalogues and newly observed data from PAU to test the performances of photo-z codes ANNZ, BPZ, ANNZ2 and DELIGHT, and we attempt to tackle the problems of missing and incomplete data. This section will answer the following research questions:

1. How do ANNZ, ANNZ2 (machine learning code), BPZ (a template code) and DELIGHT (a hybrid template-machine learning code) compare in producing photo-z for on simulated and real data in the context of the PAU Survey?

2. What kind of calibration, template choice and algorithm optimization is needed for DELIGHT to produce precise and accurate photo-z's when using a mixture of broadband and narrowband fluxes?

3. Is DELIGHT capable of dealing missing or incomplete data in the context of the PAU Survey?

Table 4: List of photo-z codes featured in this study, sorted in alphabetical order. 'TP' means template, 'EP' means empirical, while 'ML' means machine learning.

Code	Type	Source
Beagle	TP	Chevallard and Charlot (2016)
Bpz	TP	Benítez (2000)
Eazy	TP	Brammer et al. (2008)
Gazelle	TP	Kotulla and Fritze (2009)
Goodz	TP	Dahlen et al. (2010)
Hyperz	TP	Bolzonella et al. (2000)
Impz	TP	Babbedge et al. (2004)
Kcorrect	TP	Blanton and Roweis (2007)
Lephare	TP	Arnouts et al. (1999)
Lrt	TP	Assef et al. (2008)
Mizuki	TP	Tanaka (2015)
Photo-z-sql	TP	Beck et al. (2017)
Photoz	TP	Bender et al. (2001)
Rainbow	TP	Barro et al. (2011)
Satmc	TP	Johnson et al. (2013)
Spoc	TP	Finlator et al. (2007)
Wikz	TP	Wiklind et al. (2008)
Zebra	TP	Feldmann et al. (2006)
Zphot	TP	Giallongo et al. (1998)
Demp	EP	Hsieh and Yee (2014)
Annz	ML	Collister and Lahav (2004)
Annz2	ML	Sadeh et al. (2016)
Arborz	ML	Gerdes et al. (2010)
Cmnn	ML	Graham et al. (2017)
Cubanz	ML	Samui and Samui Pal (2017)
Desdm	ML	Oyaizu et al. (2008a)
Ephor	ML	Tanaka et al. (2018)
Flexzboost	ML	Izbicki et al. (2017)
Franken-z	ML	Speagle et al. (in prep.)
Gaz	ML	Hogan et al. (2015)
Gpz	ML	Almosallam et al. (2016b)
Metaphor	ML	Cavuoti et al. (2017b)
Mlpqna	ML	Brescia et al. (2014)
Photoraptor	ML	Cavuoti et al. (2015)
Skynet	ML	Graff et al. (2014)

Code	Type	Source
Somz	ML	Carrasco Kind and Brunner (2014a)
Spiderz	ML	Jones and Singal (2017)
Tpz	ML	Carrasco Kind and Brunner (2013)
Delight	TP+ML	Leistedt and Hogg (2017)
Kernelz	TP+ML	Wolf (2009)

Part II

RESEARCH WORK

4

PHOTOMETRIC AND SPECTROSCOPIC GALAXY SURVEYS

Where were you when I laid the earth's foundation? Tell me, if you understand. Who marked off its dimensions? Surely you know! Who stretched a measuring line across it? On what were its footings set, or who laid its cornerstone – while the morning stars sang together and all the angels shouted for joy?

Job 38:4-7

4.1 PHOTOMETRIC SURVEYS

In this work, galaxy photometry and morphology from various photometric surveys have been used. In the following paragraphs the sources and selection cuts of these photometry and morphology datasets will be introduced. Here we also summarise visually in Figure 13 and Figure 14 the filters and their respective responses of the various photometric surveys used in this work. The filter responses shown are from the following surveys: SDSS (Fukugita et al. 1996), CS82 (Boulade et al. 2003), COSMOS (Miyazaki et al. 2002), PAU (Martí et al. 2014), LSST (Ivezić et al. 2008) and the Stripe-82 Massive Galaxy Catalogue (S82-MGC, Bundy et al. 2012), in comparison to the original Johnson-Cousins $UBVRI$ system. The details of the photometric surveys mentioned will be discussed in the sections that follow.

4.1.1 *SDSS*

The Sloan Digital Sky Survey (SDSS, York et al. 2000) is a large scale photometric and spectroscopic survey, utilising its 2.5 m telescope (Gunn et al. 2006) at the Apache Point Observatory (APO) at New Mexico. Its primary photometric survey, the Legacy Survey of SDSS-I/II observed the northern galactic cap at high galactic latitudes between the years 2000 and 2008, imaging a sky area of over 14 555 $\deg^2$ (Aihara et al. 2011). Observing went as faint as $g \sim 23$, obtaining photometry for more than 350 million objects with a median seeing of $1.4''$ in the r-band. Photometry was observed using a large-format mosaic 120-megapixel CCD (Gunn et al. 1998), essentially simultaneously in 5 $ugriz$ bands (Fukugita et al. 1996, filter responses visualised in panel 3 of Figure 13) in a drift-scanning method known as the time-delay-and-integrate (TDI) scanning mode.

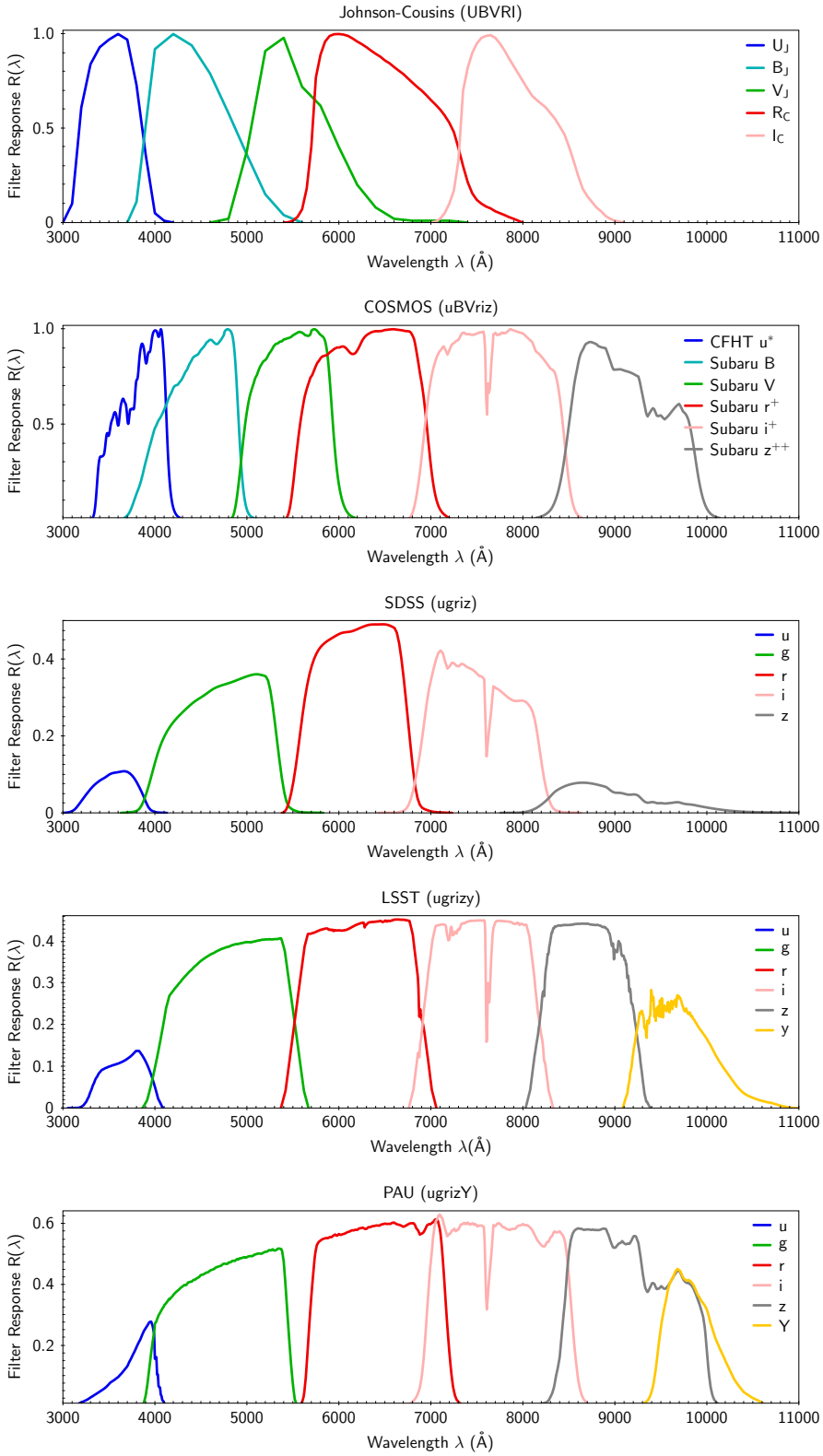

Figure 13: List of photometric filters used in this work. From top to bottom: Johnson-Cousins (*UBVRI*), COSMOS (*uBVriz*), SDSS (*ugriz*), LSST (*ugrizy*) and PAU broadbands (*ugrizY*). Note that the values of the filter responses $R(\lambda)$ are normalized differently.

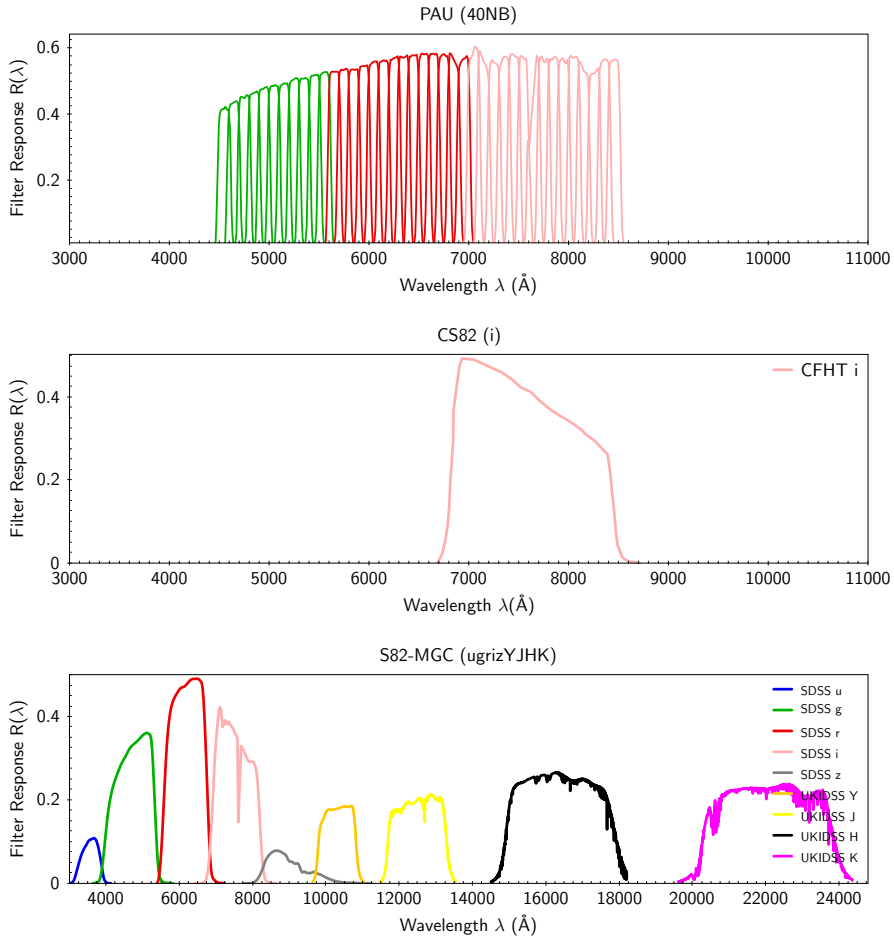

Figure 14: (*Continued from Figure 13*) List of photometric filters used in this work. From top to bottom: PAU (40NB), CS82 (single *i*-band) and S82-MGC (*ugrizYJHK*). Note the different wavelength scale for the S82-MGC.

The imaging of SDSS[1] was taken as part of the Legacy Survey, photometric data have been completed in Data Release 6 (DR6), with slight improvements of data with poor seeing in DR7 (Abazajian et al. 2009) and its final release in DR8 (Aihara et al. 2011) with improved sky subtraction, self-consistent photometric recalibration and flat-field determination. Photometry for the MGS, LRG Sample, Constant Mass (CMASS) Sample and Low Redshift (LOWZ) Sample were taken from the Legacy Survey (more on these samples in Section 4.3), however photometry and morphology were also taken from three other value-added catalogues or surveys within the SDSS database, these catalogues are the Galaxy Zoo, Stripe-82 Coadd and the Supernova Survey, which will be discussed below.

Galaxy Zoo (Lintott et al. 2008) is an internet-based project that used the collective efforts of hundreds of thousands of volunteers to classify galaxy morphologies. Galaxy shape classification is vital to study galaxy properties and formation, however due to modern galaxy surveys producing samples too large for human inspection, galaxy shapes have been classified via morphology proxies like galaxy colour, concentration index and intensity profile, which often lead to misclassifications. When Galaxy Zoo started, images of galaxies from the SDSS MGS (Strauss et al. 2002) were visually inspected by participants through a website, they were given the options to classify the shapes of these objects as either ellipticals, clockwise/anti-clockwise spirals, mergers or unknown. The results of its first data release (Lintott et al. 2011) were made public through SDSS DR8 (Aihara et al. 2011), and the classifications were found to be consistent with those of professional astronomers within 10% discrepancy, proving the 'citizen science' concept successful (Lintott et al. 2008). This first data release had been used as a training set for machine learning morphological classifications (Banerji et al. 2010), photo-z estimation (Way 2011), fundamental plane calibration (Saulder et al. 2013), and also proof that there is no evidence for a preferred spiral galaxy rotation direction in the universe (Land et al. 2008). To date, Galaxy Zoo has produced a series of 55 publications, and classification is still ongoing with new images from the Dark Energy Camera Legacy Survey (DECaLS, Schlegel et al. 2015). Classification data were based on the number of votes using a weighted system described in (Lintott et al. 2008), data can be accessed through SDSS Catalogue Archive Server[2] (CAS, Thakar et al. 2008) in the table `ZOOVOTES`, in which `P_EL` indicates the probability of an object being elliptical, `P_CS` being combined spiral. In this work we utilised the morphology classification of ellipticals and spirals in Galaxy Zoo to study the improvement of photo-z with morphology based on samples separated by their shapes. Similar to Saulder et al. (2013), we select only objects with more than 10 votes `NVOTE_TOT>10`, and classify an object as elliptical when `P_EL>0.7` and spiral when `P_CS>0.7`. These data were cross-matched with photometry and spectroscopic redshifts from the MGS and the results are shown in Section 6.2.2 and Section 6.2.3.

The *SDSS Stripe-82 Coadd* (Annis et al. 2014) is the co-addition of SDSS Stripe-82 imaging data (Jiang et al. 2014), obtained by repeated scanning along the stripe. The Stripe-82 covers an area of 227 deg^2, it is located at the equator between $-50^\circ \leq$RA$\leq 60^\circ$ and $-1.25^\circ \leq$DEC$\leq 1.25^\circ$. Among the 303 runs by SDSS in the Stripe-82 region, 84 runs were carefully selected (fields with sky brightness > 19.5 mag arcsec^{-2}, extinction < 0.2 mag etc.), these runs were co-added with an additional 39 repeated runs on the stripe, and this produced approximately 13 million

1 In this section we will focus on its photometric survey, the SDSS spectrograph and related spectroscopic surveys will be discussed in Section 4.3.1.

2 https://skyserver.sdss.org/CasJobs/.

galaxies with r-band median seeing of $1.1''$ and magnitude $15.5 < r < 23.5$, reaching on average 2 magnitudes fainter than standard SDSS observations. The co-added images have now a higher density of galaxies than before, the star-galaxy separator for this catalogue was changed to $r_{\rm PSF} - r_{\rm mod} > 0.03$ to include fainter galaxies, although stars still contaminate the galaxy catalogue by 18% (Annis et al. 2014). This wide-angle deep imaging catalogue has been used extensively in many projects, e.g. photo-z computation (Reis et al. 2012), quasar classification (Peters et al. 2015), massive galaxy evolution (Bundy et al. 2015) and even deeper co-adds (Fliri and Trujillo 2016). This co-added photometric data are stored under `RUN=106,206` in the separated Stripe-82 database on the SDSS CAS, released together with the SDSS DR7 (Abazajian et al. 2009). Photometry from the Stripe-82 Coadd were used particularly in Section 6.4, Section 6.5, Section 7.2 and Section 7.3. We chose to work with objects in the Stripe-82 region due to the availability of high quality galaxy morphology from CS82 (see Section 4.1.4) and the abundance of spectroscopic redshifts in this region, in which the photometry, morphology and spectroscopic redshifts will be cross-matched to produce training and testing samples for a large portion of this work.

Prior to Fall 2004, all observational runs on Stripe-82 were taken under photometric conditions as required for imaging in the SDSS Legacy Survey (York et al. 2000). However, runs after Fall 2005 were made as part of the *SDSS-II Supernova Survey* (Frieman et al. 2008), and observations were done with a higher cadence and often observed under poor seeing conditions ($\geq 2''$), bright sky, non-photometric conditions and low atmospheric transparency (Sako et al. 2008). This 'bad photometry', although having photometric errors at least twice as large as those from the co-add runs, was good enough to identify and measure light curves for SNIa within redshifts $0.05 < z < 0.35$, but only done after removing images taken under extremely poor photometric conditions and subjected to a special photometric calibration procedure with the coadd images as templates (Bramich et al. 2008). These photometry were released in SDSS DR7 (Abazajian et al. 2009), resulting in a photometry subset with magnitude errors larger than usual (i.e., lower $^{S}/_{N}$), especially in the u-band. In this work we are not interested in supernova detection, however we utilised this lower quality photometry to study the impact of morphology on photometric redshifts in surveys when the quality of photometry is limited. These 'bad photometry' measurements were cross-matched with high-quality CS82 morphology and spectroscopic redshifts from various surveys, and the results are shown in Section 6.5.3.

All SDSS photometric data used in this work were accessed through CAS from the SDSS DR12 (Alam et al. 2015) website, which at the start of this work was the latest data release. On top of the individual cuts mentioned above, all photometry (*ugriz* model magnitudes), morphology (r, q, f, C, p, Q and U) and their respective errors must be well defined (not -99), objects are `PRIMARY`, neither `BRIGHT`, `SATURATED` nor `BLENDED` (this applies even to photometry from the Supernova Survey).

4.1.2 *NYU-VAGC*

The New York University Value-Added Galaxy Catalogue (NYU-VAGC, Blanton et al. 2005) is a catalogue compiled with derived quantities from images and spectra of local galaxies ($z \sim 0.3$) from the SDSS DR2 (Abazajian et al. 2004). It also contained matches between major public

catalogues like the Faint Images of the Radio Sky at Twenty cm (FIRST, Becker et al. 1995), 2 Micron All Sky Survey (2MASS, Skrutskie et al. 2006), 2-degree Field (2dF) Galaxy Redshift Survey and a few others. The catalogue was created to aid the study of the local universe, it forms the basis for studies on the luminosity function, power spectra and correlation function. It was specially made to be small for easy processing on personal computers during the early 2000s. The NYU-VAGC is said to be photometrically calibrated in a more consistent way than SDSS DR2, it has an explicit description of its geometry and thus is more appropriate for studies on galaxy properties and large scale structure statistics. NYU-VAGC also derived valuable quantities for galaxies in the SDSS MGS like *K*-correction, Petrosian half-light surface brightness, and most important of all, Sérsic fit parameters (flux, index and radius). In this work we cross-matched the Sérsic index of NYU-VAGC with the MGS in order to evaluate the impact of the Sérsic index on photo-z's, as shown in Section 6.2.

4.1.3 *S82-MGC*

The Stripe-82 Massive Galaxy Catalogue (S82-MGC, Bundy et al. 2015) is a stellar mass-limited sample ($\log \frac{M_\star}{M_\odot} > 11.2$) of galaxies reaching up to redshift $z \sim 0.7$, it was constructed with the aim to study the growth history of massive galaxies. S82-MGC was built with spectroscopic redshifts from SDSS-I/II, BOSS, DEEP2 and VVDS; *ugriz* photometry from SDSS Stripe-82 Coadd; and infrared *YJHK* photometry from the UKIDSS-Large Area Survey (UKIDSS-LAS, Lawrence et al. 2007). UKIDSS-LAS utilised the Wide Field Camera on the 3.8 m United Kingdom Infrared Telescope (UKIRT) to map a large portion of the northern sky in the infrared to complement the visible range of SDSS, but the survey itself was primarily driven by the detection and discovery of brown dwarfs, young free floating planets and high-redshift quasars.

S82-MGC provided a large number of 9 photometric bands (wavelength responses visualised in panel 3 of Figure 14), the *ugriz* and *YJHK* bands were matched using synthetic aperture photometry (Synmag, Bundy et al. 2012), a very fast method which exploits galaxy profile fits in one band to predict fluxes in arbitrarily sized apertures of other bands, and this ensures consistent colours across bands from different surveys. Another important feature of S82-MGC is that it used a modified version of SDSS `TYPE=3,6` star-galaxy separator: S82-MGC set objects with $r_{\rm PSF} - r_{\rm mod} > 8$ as galaxies, $r_{\rm PSF} - r_{\rm mod} < 0.05$ as point sources, while objects in between are separated by using loci in a *J-K* v.s. *g-i* colour-colour diagram (see Bundy et al. 2015, for more details). They claimed that this lowers the contamination of stars in a galaxy sample to 1% level, compared to 10% as depicted in the SDSS Stripe-82 Photo-z Catalogue by Reis et al. (2012). A stellar mass limited subsample of the S82-MGC known as the Ukwide catalogue with the features mentioned above was created, and this sample was used in this work to study the impact of morphology on photo-z with infrared bands, and this results is shown in Section 6.4.3.

4.1.4 *CS82*

The CFHT Stripe-82 (CS82) Survey is a joint Canada-France-Brazil project. Using the MegaCam (Boulade et al. 2003) on the 3.6 m Canada-France-Hawaii Telescope (CFHT), it surveyed approximately 170 deg^2 of the equatorial Stripe-82 area. It is a relatively deep survey that maps down

to magnitude 24.1 in the CFHT/MegaPrime i-band (which is the only band in this survey, wavelength response visualised in panel 2 of Figure 14). Taking advantage of its low mean seeing of of 0.6″, a galaxy morphology catalogue with high quality morphology has been produced (Charbonnier et al. 2017, Moraes et al., in prep.) using SExtractor[3] (Bertin and Arnouts 1996) and PSFEx[4] to fit a series of de Vaucouleurs, exponential, Sérsic and bulge-disc profiles. Data from this survey has been used for several weak lensing analyses (e.g. Comparat et al. 2013; Shan et al. 2014; Hand et al. 2015; Liu et al. 2015; Battaglia et al. 2016; Li et al. 2016; Niemiec et al. 2017; Leauthaud et al. 2017; Shan et al. 2017; Pereira et al. 2018; Vitorelli et al. 2018).

The CS82 i-band magnitude and morphological parameters including the exponential radius, axis-ratio, shape probability and Sérsic index have been cross-matched with Stripe-82 photometry and spectroscopic redshifts to study the effects of high-quality morphology on photo-z's when the testing set is reweighted with respect to the CS82 Survey in Section 6.5 (see Section 3.3.4 for more details on reweighting). Morphology from this catalogue is also used in Chapter 7 to study the impact of morphology on photo-z PDFs and redshift distributions. Only objects with `SPREAD_MODEL_SER>0.008` are used, this is the star-galaxy separator used by CS82 to separate extended objects from point sources (Moraes et al., in prep.). Further cuts include `MASK=0` (not masked), `0≤FLAGS≤3` (flag for good quality source extraction), and `MAGERR_AUTO<` 0.1086 ($S/N > 10$). These cuts provide a sample of 5 977 518, where photo-z's would be estimated for in Section 7.3.

4.1.5 *PAU*

The Physics of the Accelerating Universe (PAU) Survey is a narrowband photometric galaxy survey aimed at mapping the large-scale structure of the universe up to $i \sim 23.0$. Using 6 *ugrizY* broadbands and 40 narrowbands of FWHM 125 Å (filter response visualised in Figure 13 and Figure 14), PAU aims to achieve quasi-spectroscopic redshifts with a resolution of $R_S \sim 50$ with precision of $\sigma_{\rm RMS} < 0.0035(1+z)$ for galaxies with $i < 22.5$. PAU uses the PAUCam (Castander et al. 2012) on the 4 m William Herschel Telescope (WHT) at Observatorio del Roque de los Muchachos (ORM) in La Palma, it has observed more than 50 deg^2 of sky since the beginning of 2016, and observations are still ongoing. The PAU forced-aperture coadded photometry has its aperture defined by elliptical sizes of COSMOS morphology, therefore the fluxes measured are based on an aperture radius covering a fixed percentage of light. Martí et al. (2014) has conducted a photo-z performance test in the context of PAU using the template-based code BPZ on a mock galaxy catalogue (Section 4.2.2).

In this work we used the early data release of PAU with `PRODUCTION_ID=697` (using circular aperture with 62.5% light radius), photometry with $i_{\rm auto} \leq 22.5$, `N_BANDS=40` (detected in 40 bands, i.e. no missing bands) and `TYPE=0` (extended objects) were used. Cross-matching with spectroscopic redshifts from zCOSMOS yields a sample of 8100 objects used for training and testing respectively. In this work we test the photo-z performance of real PAU data using ANNz2 and DELIGHT, to test the robustness of these codes when dealing with more than 10 inputs, and to see their ability to deal with missing data. The results of these tests are shown in Section 8.3.

3 https://www.astromatic.net/software/sextractor
4 https://www.astromatic.net/software/psfex

4.1.6 *COSMOS*

The Cosmic Evolution Survey (COSMOS, Scoville et al. 2007) was designed to probe the correlated evolution of galaxies, star formation, dark matter and AGNs with large-scale structure within the redshift range $0.5 < z < 6.0$. The survey covers a sky area of 2 deg^2 and is known for its high sensitivity and depth. The COSMOS field has an exceptionally low and uniform Galactic extinction ($E_{B\text{-}V} \sim 0.02$), and therefore allowed the production of high quality photometry, spectroscopic redshifts and photo-z's (Ilbert et al. 2009).

In this work we used photometry from the COSMOS2015 Catalogue (Laigle et al. 2016). It is a highly complete mass-selected sample to very high redshifts, highly optimised for the study of galaxy evolution and environments in the early universe. The COSMOS2015 Catalogue provides 30 band photometry ranging from X-ray to radio wavelengths, all these have been observed through multiple facilities including GALEX (Martin et al. 2005), CFHT, Subaru Telescope (Miyazaki et al. 2002), Ultra Visible and Infrared Survey Telescope for Astronomy (UltraVISTA, McCracken et al. 2012), and Spitzer Large Area Survey with HSC (SPLASH, Steinhardt et al. 2014). From this catalogue we only use the CFHT u-band (Boulade et al. 2003) and Subaru $BVriz$ bands (Miyazaki et al. 2002) to match the narrow band photometry of PAU, since at present PAU has not observed broadband photometry. The photo-z performance of PAU narrowbands are compared to those of COSMOS broadbands, and the results are shown in Section 8.3.

4.2 SIMULATED GALAXY CATALOGUES

Galaxy simulations allow us to follow the growth of cosmological structure, model non-linear effects in clustering and study galaxy formations, lensing and redshift distortions. Galaxy catalogues produced from these simulations play an important role in characterising telescopes, setting up goals of surveys and providing preparatory ground work before real telescope observations take place. Each simulation differs from one another by how the galaxies are generated. In this work we used 3 mock / simulated galaxy catalogues, one of which is the Buzzard Simulation in the context of the 6.5 m LSST (Ivezić et al. 2008), a next-generation upcoming ground-based survey designed to map the sky at an unprecedented depth and speed. The other 2 mock catalogues are for the context of PAU, they are used to calibrate and test our photo-z codes in dealing with large number of inputs. The following paragraphs summarise the properties of the simulated galaxy catalogues used in this work.

4.2.1 *Buzzard Simulation*

The BUZZARD-HIGHRES-V1.1 Catalogue (DeRose et al., in prep., but also described in Hoyle et al. 2017) is a mock first-year DES Catalogue constructed using a dark matter only simulation. This N-body simulation contained 2048^3 particles in a volume of $0.064h^{-3}$ Gpc^3. Dark matter halos were identified using ROCKSTAR (Behroozi et al. 2013) and populated with galaxies with a stellar mass $M_\star$ and absolute r-band magnitude M_r determined using a sub-halo abundance matching model, constrained to match both projected two-point galaxy clustering statistics and an observed conditional stellar mass function (Reddick et al. 2013). SEDs are assigned to each

galaxy using Adding Density Dependent Spectral Energy Distributions[5] (ADDSEDS, DeRose et. al., in prep.) by training an empirical relation between absolute r-band magnitude, local galaxy density and SED using a sample of 10 000 SDSS galaxy spectra from the NYU-VAGC. For each simulated galaxy, a random galaxy from the training set with similar M_r and local galaxy density was chosen and its SED then assigned to the simulated galaxy. Given the SED, M_r and redshift, the LSST *ugrizy* apparent magnitudes were computed. Magnitude errors were assigned using the simple model described in Ivezić et al. (2008), assuming full 10-year depth observations had been completed but with minor tweaks. As the simulation is catalogue-based, no image level effects such as object blending, sky background contamination and lensing magnification were included, and neither were stars and AGNs considered in the sample.

The total catalogue covered a sky area of 400 deg^2 and contained 238 million galaxies up to apparent magnitude $r \sim 29.0$ and redshift $z \leq 8.7$, but for the purpose of the Data Challenge 1 conducted by the LSST-DESC PZ, data from only ~ 8 deg^2 were used, and the redshift was cut at $z < 2$ due to systematic problems with galaxy colours at $z > 2$. To allow template-based photo-z codes to run, 150 ordered templates were produced by using a K-means clustering algorithm on the SDSS SEDs. This essentially partitions the 10 000 SDSS SEDs with coefficients $\{z, M_r\}$ into 150 means in the coefficient space, providing a reasonable approximation for a spanning SED set. The sample was split randomly to produce 111 171 galaxies for training and 1 000 883 for testing. Analysis is restricted to the Gold Sample: a cut with $i < 25.3$, giving $S/N \sim 30$ for most galaxies, cutting the final testing set to 399 356 galaxies. Results based on our part in this collaboration work will be discussed in Section 7.4.

4.2.2 *PAU Mock Catalogue*

A mock catalogue was generated by Martí et al. (2014) to study the photo-z performance of PAU with respect to the choice of number of filters and their widths. About a million galaxies were generated with observed magnitudes $m_b \pm \sigma_{m_b}$ for each of the 6 broadbands and 40 narrowbands, for galaxies at redshift z with template (spectral type) t. The catalogue was generated using the cumulative luminosity function. Knowing that the absolute magnitude M is dependent on redshift z and spectral type t, Equation 38 can be written as

$$N(x) = \int_{-\infty}^{M_{\text{lim}}(z,t)} \phi_\star x^\alpha e^{-x}\, dx. \tag{72}$$

As limiting magnitudes are usually set in apparent magnitudes, M_{lim} can be related to m_{lim} via the equation

$$m = M + 5 \log d_L(z) + 25 + K(z,t), \tag{73}$$

where d_{L} is the luminosity distance (Equation 25) and $K(z,t)$ the K-correction (Equation 48).

The galaxy catalogue is simulated by setting the redshift range to $0 < z < 6$, limiting magnitude of $m_{\text{lim}} = 24.0$ in the SDSS r-band and spectral types using the 66 LEPHARE templates (see Section 5.6.3). Its redshift distribution is based on GOODS. The values of $M_\star$, $\phi_\star$ and α were

5 https://github.com/vipasu/addseds

modelled to be redshift dependent such that $\{M_\star, \log\phi_\star, \alpha\} \propto e^{-(1+z)^a}$, such that their values match those of Dahlen et al. (2005) in the B-band. With every galaxy characterised by three parameters $\{z, t, M\}$ following the luminosity function, the apparent magnitude for the reference band m_0 (SDSS r-band) is calculated, and other bands m_b are computed in reference to m_0 via

$$m_b = m_0 + 2.5 \log \left[\frac{\int_0^\infty f_t((1+z)\lambda) R_0(\lambda) \lambda \, d\lambda}{\int_0^\infty f_t((1+z)\lambda) R_b(\lambda) \lambda \, d\lambda} \frac{\int_0^\infty R_b(\lambda) \lambda \, \frac{d\lambda}{\lambda}}{\int_0^\infty R_0(\lambda) \lambda \, \frac{d\lambda}{\lambda}} \right] . \tag{74}$$

Note that the PAU filters used to generate this mock catalogue were later updated, though having slightly different filter responses, the FWHM of each band remain the same. Noise is then added to the magnitudes based on Equation 44 multiplying with a normal variable (Gaussian noise), and limiting the $S/N \leq 50$ to account for possible photometric calibration issues (Martí et al. 2014). The sky brightness spectrum of La Palma, Spain was provided by Benn and Ellison (1998) and included in this simulation.

We made a selection cut of $i < 22.5$ analogous to the bright sample in Martí et al. (2014), yielding a sample of 244 737 objects, this sample was used to test the performance of ANNZ2 and DELIGHT. The results are shown in Section 8.2.

4.2.3 *MICE Simulation*

The *Marenostrum Institut de Ciències de l'Espai* (MICE) Simulation (Crocce et al. 2010) is a large N-body simulation carried out on the Marenostrum supercomputer, and it provided a simulation with large cosmological volume with good mass resolution. MICE aimed to study structure formation and evolution at very large scales, to accurately model and calibrate basic cosmological probes which will be used by upcoming large astronomical surveys. The simulation ran with the GADGET-2 code (Springel 2005) assuming a ΛCDM universe with $\{\Omega_{m,0}, \Omega_{\Lambda,0}, \Omega_{b,0}, \sigma_8, h\} = \{0.25, 0.75, 0.044, 0.8, 0.7\}$. One of its latest simulations was the MICE Grand Challenge (Fosalba et al. 2015) run, simulating 4096^3 particles in a comoving volume of $27h^{-3}$ Gpc^3, spanning 5 orders of magnitude in dynamical range. This simulation was used to build a halo catalogue, and from it a lightcone galaxy mock (MICECAT) was produced using a hybrid technique that combines halo occupation distribution and halo abundance matching. The catalogue construction process is beyond the scope of this paper, but the reader could refer to Crocce et al. (2015) and Carretero et al. (2015) for further details. The latest publicly available galaxy mock produced (MICECAT V2) has a sky area of 5000 deg^2, it is complete to $i < 24$ and with redshift reaching $z <\sim 1.4$.

In this work the PAU-MICE2 Simulated Sample was constructed, where fluxes for PAU narrowband filters have been estimated through interpolation, taking the true redshift, COSMOS SEDs (Ilbert et al. 2009), extinction curve and its normalisation as inputs. Version 2 of the catalogue which does not include emission line fluxes was used, and this catalogue contained about 1 million galaxies, downloaded from CosmoHub[6] (Carretero et al. 2018). This sample is used to calibrate the settings and choice of hyperparameters of DELIGHT. The results for this sample is shown in Section 8.2.

6 http://cosmohub.pic.es/

This work uses galaxy redshifts from at least 7 different surveys, either independently or combined in order to test the performance of photo-z's. More specifically, these samples were mainly used as training and testing sets for machine learning photometric redshift methods (more on this in Chapter 5), and the spectroscopic redshifts act as the true redshift in comparison to the photometric redshifts. This list of spectroscopic samples is essentially assumed to be a complete list of usable spectra and in a certain way they constrain our analysis of photometric redshifts. In this section we introduce all spectroscopic galaxy surveys in which spectroscopic redshifts were used in this work, including their sample selection criteria, and a brief mention of the section in which the data are used.

4.3.1 *SDSS*

The photometric aspects of the Sloan Digital Sky Survey (SDSS) have been discussed in Section 4.1.1, here we focus on the spectroscopic aspects of the survey. The SDSS spectrograph (Smee et al. 2013) was used in the SDSS-I/II Legacy Survey (York et al. 2000) from year 2000-2008 which obtained spectroscopic redshifts for about 930 000 galaxies as faint as $r_{\mathrm{Pet}} \sim 17.77$ and about 120 000 quasars up to $i_{\mathrm{PSF}} \sim 19.1$. The SDSS spectrograph is a 640-fibre-fed multi-object spectrograph covering a wavelength range of 3800 Å to 9200 Å with a spectral resolution of $R_{\mathrm{S}} \sim 2000$ (Smee et al. 2013). Spectroscopic target selection was conducted after imaging data were reduced and calibrated. In the following paragraphs we list the three main spectroscopic samples used in this work.

The first sample is the *Main Galaxy Sample* (MGS, Strauss et al. 2002). The MGS is a uniform and complete low-redshift galaxy sample ideal for studies of large-scale-structure and characteristics of galaxy populations in the local universe. The median redshift of this sample is $z \sim 0.1$, and the target selection is as follows: a tighter star-galaxy separation $r_{\mathrm{PSF}} - r_{\mathrm{mod}} \geq 0.3$ (originally ≥ 0.145); `DETECTED` in r (magnitude more than 5σ above the sky after smoothing with a PSF filter), not `SATURATED` (object containing more than one saturated pixel); not `BRIGHT` (objects in initial bright object-finding sequence); not `BLENDED` (object brightness with multiple peak); and $r_{\mathrm{Pet}} \leq 17.77$. Further cuts based on their surface brightnesses (based on detection limits of imaging data) were made and the reader could refer to Strauss et al. (2002) for more details. This sample has been used by Zehavi et al. (2002) to study galaxy clustering and Saulder et al. (2013) to calibrate the fundamental plane. Spectroscopic redshifts from the MGS have been obtained via the SDSS CAS by setting the flag `TARGET=MAIN`, with `ZWARNING=0` to ensure good quality redshifts. These redshifts were cross-matched with SDSS photometry and morphology to yield a sample of 726 525 galaxies, and is used in Section 6.2. The sample is further cross-matched with data from the *Galaxy Zoo* (Lintott et al. 2011, see Section 4.1.1) to obtain more accurate shape measurements, yielding a sample of 634 101 galaxies. The results of this study will be discussed in Section 6.2.2.

The second sample is the *Luminous Red Galaxy Sample* (LRG, Eisenstein et al. 2001). This is a colour and magnitude selected sample, a group of luminous intrinsically red galaxies that extends fainter and farther than the MGS, observable despite lower S/N due to the presence of the strong 4000 Å break. This sample is suited to study large-scale structures, clusters and the

Table 5: Photometric cuts for the LRG Sample as stated in Eisenstein et al. (2001).

Cut I ($z < 0.4$)	**Cut II** ($z \geq 0.4$)
$r_{\rm Pet} < 13.1 + 3.333 c_{\parallel}$	-
$r_{\rm Pet} < 19.2$	$r_{\rm Pet} < 19.5$
$\lvert c_{\perp} \rvert < 0.2$	$c_{\perp} > 0.45 - 0.167(g\text{-}r)$
-	$g\text{-}r > 1.30 + 0.25(r\text{-}i)$
$r_{\rm PSF} - r_{\rm mod} > 0.3$	$r_{\rm PSF} - r_{\rm mod} > 0.5$

evolution of giant elliptical galaxies up to $z \sim 0.5$, and it provided the first observation of BAO (Eisenstein et al. 2005).

As the 4000 Å break moves between the SDSS g and r-bands at redshift of $z \sim 0.4$, different definitions were used as to whether an object is considered an LRG. A rotated coordinate system in colour space was defined as follows (Eisenstein et al. 2001):

$$c_{\perp} = -0.25(g\text{-}r) + (r\text{-}i) - 0.18, \tag{75}$$

$$c_{\parallel} = 0.7(g\text{-}r) + 1.2(r\text{-}i) - 0.216, \tag{76}$$

where $c_{\perp} = c_{\parallel} = 0$ marked the centre of the observed locus in the g-r versus r-i plane. The target selection for cuts above and below redshift $z = 0.4$ (defined as Cut II and Cut I respectively) are summarised in Table 5. On top of that, both cuts have $\mu_{{\rm Pet},r} < 24.2$ mag arcsec^{-2} (removing objects with low surface brightness), and objects are neither `SATURATED` nor `BRIGHT` in gri. Redshifts from the LRG Sample were obtained via CAS, using the flag `TARGET=GALAXY_RED` and setting `ZWARNING=0` to remove low quality redshifts. Cross-matching with SDSS photometry and morphology we obtain a sample of 708 650 galaxies, and this sample is used in Section 6.3.

The third sample is the *Quasar Sample* (Richards et al. 2002). The SDSS Quasar Sample existed to study the evolution of the quasar luminosity function and the spatial clustering of quasar as a function of redshift. (Richards et al. 2002). It was meant to increase the number of quasars with redshifts $0 < z < 5$ by a factor of hundred to complement the Large Bright Quasar Survey (LBQS, Hewett et al. 1995) and the 2dF-SDSS LRG and QSO (2SLAQ) Survey (Croom et al. 2009). The quasar target selection is too complex to be discussed here, the reader could refer to Figure 1 of Richards et al. (2002) for more details. But generally, the quasars were selected to be $i_{\rm PSF} > 15$ to exclude bright objects; neither `BRIGHT`, `SATURATED`, `BLENDED` nor `EDGE` (lying close to edge of frame); includes objects of `TYPE=3` (extended objects) and `TYPE=6` (point sources); and goes through a series of PSF colour and radio magnitude cuts from FIRST. As we are mainly interested in extended quasars or AGNs (`TYPE=3`), a small sample of 7879 `TYPE=3` quasar redshifts were used in Section 6.4.2 to study the improvement of morphology on a sample with an imperfect star-galaxy separator. 5380 `TYPE=6` quasar redshifts were also used in the same section, to test if the converse is true.

Table 6: Photometric cuts for the LOWZ and CMASS Sample as stated in Dawson et al. (2013).

LOWZ ($0.15 < z < 0.43$)	**CMASS** ($0.43 < z < 0.70$)	**Remarks**
$\lvert c_\perp \rvert < 0.2$	$d_\perp > 0.55$	Colour boundary
$r_{\rm cmod} < 13.5 + 3.33 c_\parallel$	$i_{\rm cmod} < 18.58 + 1.6 d_\perp$	Sliding magnitude cut
$16 < r_{\rm cmod} < 19.6$	$17.5 < i_{\rm cmod} < 19.9$	Faint and bright limits
-	$r_{\rm mod} - i_{\rm mod} < 2$	Removes colour outlier
-	$i_{\rm fib2} < 21.5$	Ensure success rate
$r_{\rm PSF} - r_{\rm cmod} > 0.3$	$i_{\rm PSF} - i_{\rm mod} > 0.2 + 0.2(20.0 - i_{\rm mod})$	Star-galaxy separation
-	$z_{\rm PSF} - z_{\rm mod} > 9.125 - 0.46 z_{\rm mod}$	

4.3.2 *BOSS*

The Baryon Oscillation Spectroscopic Survey (BOSS, Dawson et al. 2013) is part of the SDSS-III (Eisenstein et al. 2011). It was the largest of the 4 surveys observing from 2008-2014. BOSS aimed to measure the cosmic expansion history of the universe up to percent-level precision by mapping an immense volume of sky, its goal was to obtain the spatial distributions of galaxies and quasars to obtain the characteristic scale imprinted by the BAO in the early universe. The BOSS spectrograph (Smee et al. 2013) was an upgrade to the original SDSS spectrograph: it has an increased detector size which extended the wavelength range to between 3600 Å and 10 400 Å; fibres were increased from 640 to 1000 per exposure; optical throughput was increased, achieving at least 1 mag deeper than SDSS; and it achieved a spectral resolution of $R_{\rm S} > 1000$. The improved spectrograph observed a spectroscopic footprint of 10 000 deg^2, allowing the measurements of 1.5 million galaxy redshifts up to $z \sim 0.7$ and the detection of the Lyman-α forest for quasars at redshift $2.2 < z < 3.0$. Target selections for LRGs have been extended to reach $i \sim 19.9$, and quasars to $g \sim 22$. The way spectra are classified has changed too: this time spectra are compared to a full range of 'eigentemplates' based on principle component analysis (PCA) of multiple galaxies, quasars and star spectra. Similar to the SDSS Legacy Survey, BOSS redshifts are divided into three different samples, each discussed in the following paragraphs. All data below were taken from SDSS CAS DR12 (Alam et al. 2015).

The first sample is the *Low Redshift Sample* (LOWZ), with galaxies having a redshift range of $0.15 < z < 0.43$. This sample uses a set of colour-magnitude cuts that follow the predicted evolution of a passively evolving stellar population with redshift. The selected galaxies are the brightest and reddest of the low-redshift galaxy population. The targeting cuts are similar to those of SDSS LRGs but includes fainter galaxies, it is summarised in the first column of Table 6. LOWZ galaxies are selected using the flags `BOSS_TARGET1=1`, `SPECPRIMARY=1` and `ZWARNING_NOQSO=0`. Redshifts of the LOWZ Sample were cross-matched with SDSS photometry to produce a sample of 327 637 galaxies used in Section 6.3.

The second sample is the *Constant Mass Sample* (CMASS), covering a redshift range of $0.43 < z < 0.7$. Although similar to the SDSS LRG Sample, CMASS galaxies are not restricted to red galaxies but instead makes a stellar mass-limited sample of objects of all intrinsic colours. The luminosity cut is also lowered to allow a wider colour range to provide a less biased sample for massive galaxy evolution studies. The selection cuts are summarised in Table 6 in comparison with the cuts for the LOWZ Sample, where the auxiliary colour $d_\perp = -0.125(g\text{-}r) + (r\text{-}i)$ and

i_{fib2} is the 2″ fibre magnitude. The cuts are similar to SDSS LRG Cut II but extended to bluer galaxies. The CMASS selection also introduced a sliding colour-magnitude cut to select more massive objects uniformly as a function of redshift. The CMASS galaxies are selected from CAS with flags `BOSS_TARGET1=2`, `SPECPRIMARY=1`, `ZWARNING_NOQSO=0` and `CHUNK` not from `BOSS1` and `BOSS2`. CMASS redshifts were cross-matched with SDSS photometry and morphology to produce a sample of 884 424 galaxies to be used in Section 6.3.

The third sample is the BOSS *Quasar Sample*, where its primary goal is to map the large-scale distribution of neutral hydrogen via absorption in Lyman-α forests. The target selection for the BOSS Quasar Sample used multiple approaches including kernel density estimation and artificial neural networks, in which the reader could refer to Dawson et al. (2013) for more details. Similar to the SDSS Quasar Sample, we are more interested in extended quasars and AGNs, therefore selecting objects with `TYPE=3` and `CLASS=QSO`, to be used in the training set to produce photo-z's for the CS82 Photo-z Catalogue (Section 7.3). Only 552 quasar redshifts were obtained from this sample.

4.3.3 *2SLAQ*

The 2-degree Field (2dF) Galaxy Redshift Survey (Colless et al. 2001) aimed to reveal large scale structures and the clustering of matter when the Universe was at $z \sim 0.6$ in order to understand the evolution of massive galaxies. This survey used the 2dF spectrograph mounted at the primary focus of the 3.9 m Anglo-Australian Telescope (AAT) at the Anglo-Australian Observatory (AAO). The observations went on between 1997 and 2002, surveyed both the north and south galactic poles and covered an area of about 1500 deg^2. The 2dF has a resolution of $R_S \sim 1000$, with wavelength range 5050 Å to 7250 Å, able to measure about 400 spectra in one go. As a result of the survey, the 2dF-SDSS LRG and QSO Catalogue (2SLAQ, Cannon et al. 2006) was created, it has about 15 000 LRGs brighter than $i = 19.8$ and having redshifts between $0.45 < z < 0.8$. The LRGs were selected using Cut II of the SDSS LRG selection (Eisenstein et al. 2001), as shown in Section 4.3.1.

Redshifts with quality `Q=3,4,5` from the 2SLAQ Catalogue were selected for the initial calibration of hyperparameters for the photo-z code ANNZ, and also used in Section 6.3. The sample is cross-matched with SDSS DR12 photometry, yielding a sample of 9718 galaxies.

4.3.4 *DEEP2*

The DEEP2 Galaxy Redshift Survey (Newman et al. 2013) was designed to conduct a comprehensive study on the properties of high redshift massive galaxies and its environment. DEEP2 is primarily sensitive to galaxies of redshift $z < 1.45$, the limit where the O II 3727 Å doublet moves beyond the red limit of its typical spectral coverage. DEEP2 used the Deep Imaging Multi-Object Spectrograph (DEIMOS Faber et al. 2003) mounted on the 10 m Keck II telescope in Hawaii to observe 2.8 deg^2 of sky, it is a survey which reaches a depth of $r \sim 24.1$. With a very high spectral resolution of $R_S \sim 6000$, it yields 38 000 reliable redshift measurements which produced numerous science findings well summarised in Newman et al. (2013). The objects are pre-selected using

BRI photometry ($18.5 < R < 24.1$) from the CFHT, using colour cuts to remove redshifts $z < 0.7$ to obtain a sample with mean redshift $z \sim 1$.

The DEEP2 survey has an overlap area of 0.5 deg^2 with the Stripe-82 region, and objects in this area are cross-matched with SDSS Stripe-82 photometry and CS82 morphology to yield a galaxy sample of 11 858 objects to be used throughout Chapter 6 and Chapter 7. Redshifts from this sample are obtained from the DEEP2 DR4 Redshift Catalogue[7], keeping only objects with `ZQUALITY`$\geq$`3`.

4.3.5 *VVDS*

The VIMOS VLT Deep Survey (VVDS, Le Fèvre et al. 2005) is a deep representative galaxy survey, it aimed to trace the large-scale distribution of galaxies on comoving scales reaching $100h^{-1}$ Mpc, making it suitable for studying the bright end of the luminosity function and the massive end of the mass function up to $z \sim 1$. VVDS used the Visible Multi-Object Spectrograph (VIMOS, Le Fèvre et al. 2003) at the European Southern Observatory's Very Large Telescope (ESO-VLT), currently mounted on the 8.2 m Melipal telescope (UT3). it is a multi-purpose wide-field survey instrument used for direct imaging, multi-slit spectroscopy and integral field spectroscopy, it can observe up to 600 objects simultaneously and attain a spectral resolution of up to $R_S \sim 2500$ (Le Fèvre et al. 2005). VVDS is a magnitude-limited survey, targets are selected based on cuts on the CFHT *i*-band Kron magnitude. Stars were also not removed a priori using colour or compactness criteria to ensure no biases against compact galaxies and AGNs are made. VVDS has provided an unprecedented number of 933 galaxies with spectroscopic redshifts in the redshift desert $1.5 < z < 2.5$ (Le Fèvre et al. 2013).

VVDS has three main surveys: the Wide ($i < 22.5$, $z \sim 1$), Deep ($i < 24$, $0 < z < 5$) and Ultra-Deep ($i < 25$) surveys. In this work we only used redshifts from the VVDS-Wide (Garilli et al. 2008) survey, it has an overlapping area of 3.6 deg^2 within the Stripe-82 region, with a limiting magnitude of $i \sim 22.5$. Objects from the VVDS-Wide (VVDS-F2217+00) Sample[8] with `ZFLAG=3,4` (galaxies) and `ZFLAG=13,14` (quasars) are cross-matched with SDSS Stripe-82 photometry and CS82 morphology, contributing 3949 redshifts to be used throughout Chapter 6 and Chapter 7.

4.3.6 *WiggleZ*

The WiggleZ Dark Energy Survey (Drinkwater et al. 2010) is a survey particularly focused on observing ELGs, its aim was to measure the precise scale of the BAO imprinted on the spatial distribution of galaxies at look-back times between 4-8 Gyrs and measure the change of $w(z)$ to test various dark energy models. WiggleZ used the AAOmega spectrograph (Sharp et al. 2006) on the 3.9 m AAT in Australia, it has a spectral resolution of $R_S \sim 1300$ and an observable wavelength range between 3700 Å and 9500 Å, to date it has observed more than 240 000 redshifts within $0.2 < z < 1.0$ over a sky area of 1000 deg^2. Galaxies are targeted using ultraviolet photometry from GALEX to select blue and bright star-forming galaxies, when compared to SDSS, these galaxies have less significant non-linear structures and yield a lower clustering amplitude. The

7 http://deep.ps.uci.edu/DR4/zcatalog.html
8 http://cesam.lam.fr/vvds/vvds_download.php

target selection is as follows: a flux limit of $NUV < 22.8$; detected optically at $20.0 < r < 22.5$ in SDSS and the Red-sequence Cluster Survey 2 (RCS-2, Gilbank et al. 2011); a colour cut of $FUV - NUV \geq 1$ (far ultraviolet and near ultraviolet bands) to select high redshift galaxies (Lyman break enters the FUV filter at $z \sim 0.5$); $-0.5 \leq NUV - r \leq 2$ to select blue star-forming galaxies; and other low redshift rejection criteria further discussed in Drinkwater et al. (2010). The selected galaxies and redshift range made WiggleZ the first BAO survey to sample the transitional matter to Λ-dominated epoch, it obtained BAO measurements with an aggregate distance precision of 3.8% at a central redshift $z = 0.6$.

WiggleZ has a 45.3 deg^2 survey area overlapped with the Stripe-82 region. Galaxy redshifts from the WiggleZ DR1 database[9] with quality flag `QOP>3` were cross-matched with SDSS Stripe-82 photometry and CS82 morphology to provide a sample of 7648 redshifts to be used throughout Chapter 6 and Chapter 7.

4.3.7 *zCOSMOS*

The zCOSMOS Survey (Lilly et al. 2007) was designed to characterize the environments of COSMOS galaxies between 100 kpc and 100 Mpc scales to produce diagnostic information on galaxies and AGNs. zCOSMOS observes the COSMOS field, and similar to VVDS it used the VIMOS spectrograph as well. There are two main surveys in zCOSMOS: zCOSMOS-Bright observed 20 000 galaxies in a sky area of 1.7 deg^2, these galaxies have magnitudes $15 < I_{\text{auto}} < 22.5$ and redshift $0.1 < z < 1.2$, its spectral range is in the red (rest-frame wavelength 5550 Å to 9650 Å) to follow strong spectral features around the 4000 Å break to as high redshifts as possible; while zCOSMOS-Deep observed 10 000 galaxies in a sky area of 1.0 deg^2, these galaxies are colour selected in BzK and UGR (see Daddi et al. 2004; Steidel et al. 2004, for more details) and redshift $1.4 < z < 3.0$, its spectral range is in the blue (rest-frame wavelength 1200 Å to 1700 Å) to pick up stronger absorption features.

In this work we used data from zCOSMOS-Bright DR3[10], galaxies with redshift confidence class 3 to 5 were selected and cross-matched with PAU photometry to produce a sample of 19 169 galaxies. This sample is used to test the performance of the photometric redshift code Delight (see Section 5.5) in a narrowband survey in Chapter 8.

4.3.8 *Others*

In this study, redshifts from 3 other surveys have been studied and used for Section 6.5, however they are not included in the results due to various reasons. In the following points we summarise these surveys and the reasons for not using them.

- *The Prism Multi-object Survey* (PRIMUS, Coil et al. 2011). PRIMUS is a low spectral resolution ($R_S \sim 40$) spectroscopic survey which traded higher-resolution redshifts for a faster way to sample over a wide area. It covered 9.1 deg^2 of sky including fields of GALEX, Spitzer[11] and Chandra[12], to a depth of $i \sim 23.5$, providing 130 000 redshifts up to $z \sim 1.2$

9 http://wigglez.swin.edu.au/site/data.html
10 http://www.eso.org/qi/catalog/show/65
11 http://www.spitzer.caltech.edu/
12 http://chandra.harvard.edu/

for galaxies and $z \sim 5$ for AGNs. PRIMUS used a low-dispersion prism and slitmask in the Inamori Magellan Areal Camera and Spectrograph (IMACS, Bigelow and Dressler 2003), mounted on the Magellan I Baade 6.5 m telescope at the Las Campanas Observatory, Chile. The choice of prism over transmission grating decreases the spectral resolution by a lot which however increases the volume of spectroscopic redshifts obtained. Redshifts from this survey were not used as Bonnett et al. (2016) showed the existence of large percentage differences in redshift between PRIMUS and other surveys. The selection effects with the capping of redshift at $z = 1.2$ may affect our results since we use redshifts up to $z \sim 1.5$.

- *The Second Canadian Network for Observational Cosmology Redshift Survey* (CNOC2, Yee et al. 2000). CNOC2 covers four patches of sky totalling 1.5 deg^2, it was the first large redshift survey of faint field galaxies to investigate the evolution of clustering. CNOC2 used the Multi-Object Spectrograph / Subarcsecond Imaging Spectrograph (MOS/SIS, Le Fèvre et al. 1994) mounted on the 3.6 m CFHT, observing galaxies up to $R \sim 21.5$, producing a sample of 6000 redshifts within redshift $0.1 < z < 0.6$. Redshifts from this survey were not used as the number of objects cross-matched to SDSS Stripe-82 photometry were too insignificant.

- *The VIMOS Public Extragalactic Redshift Survey* (VIPERS, Guzzo et al. 2014). VIPERS aimed to improve the key statistical descriptions of current galaxy distributions, study clustering and redshift space distortion at an epoch when the universe was roughly half its current age. VIPERS used the VIMOS spectrograph at ESO-VLT and was the first VIMOS redshift survey where its data reduction was performed with a fully automated pipeline. The survey covered 23.5 deg^2 sky area within the CFHT-Wide field and produced a sample of 86 775 galaxies with magnitude $i < 22.5$ and redshift $0.5 < z < 1.5$. Similar to the reason for CNOC2, the number of redshifts cross-matched with SDSS Stripe-82 photometry was too small to make an impact on the results and was therefore not included in the analysis.

5

PHOTOMETRIC REDSHIFT ALGORITHMS AND METRICS

But each one must examine his own work, and then he will have reason for boasting in regard to himself alone, and not in regard to another.

Galatians 6:4

5.1 INTRODUCTION

Before we dive into the results of this work, we first need to understand the tools and methods used to conduct this study. In this chapter, the 4 main photo-z codes used throughout this work, namely ANNZ (Section 5.2), ANNZ2 (Section 5.3), BPZ (Section 5.4) and DELIGHT (Section 5.5) will be introduced; their functionalities and algorithms will be explained in detail. As BPZ and DELIGHT rely on the use of templates, we dedicate Section 5.6 to briefly describe the template sets used in this work. As multiple performance metrics were used to evaluate photo-z's throughout this work, we define the various performance metrics for point estimates, PDFs and distributions used in Section 5.7.

This thesis is a collection of work using many different codes, on different photometric and spectroscopic samples, thus it may be difficult to navigate through the thesis without a map. Here we provide the reader with Table 7 to easily navigate around this document, and also providing the reader a bird's eye view of this work.

5.2 ANNZ

ANNZ is an ANN redshift estimation library introduced by Collister and Lahav (2004), it is one of the earliest machine learning photo-z codes introduced to the astronomy community. ANNZ has been used by many for photo-z estimation (Collister et al. 2007; Bonfield et al. 2010), and concurrently with other photo-z codes (Hildebrandt et al. 2010; Abdalla et al. 2011; Reis et al. 2012; Sánchez et al. 2014; Bundy et al. 2015; Soo et al. 2018). It has also been employed for morphological classification of galaxies (Lahav et al. 1996; Ball et al. 2004; Banerji et al. 2010) and star-galaxy separation (Bertin and Arnouts 1996; Soumagnac et al. 2015; Kim and Brunner 2017). ANNZ is publicly downloadable from `http://www.homepages.ucl.ac.uk/~ucapola/annz.html`.

Table 7: A summary of the photometry, morphology, spectroscopy and photometric redshift algorithms used in different sections and pages in this work. The different photometric surveys mentioned below are discussed in Section 4.1, the simulated galaxy catalogues in Section 4.2 and the spectroscopic samples in Section 4.3.

Sec.	Pg.	Short Description	Photometry / Morphology	Spectroscopy	Codes Used
6.2.1	146	Photo-z + morphology on the MGS	SDSS, NYU-VAGC	SDSS	ANNZ
6.2.2	151	Photo-z + morphology on ellipticals v.s. spirals	SDSS, NYU-VAGC, Galaxy Zoo	SDSS	ANNZ
6.2.3	154	Photo-z + fundamental plane	SDSS, NYU-VAGC, Galaxy Zoo	SDSS	ANNZ
6.3	160	Photo-z + morphology on colour selected samples	SDSS	2SLAQ, SDSS, BOSS	ANNZ
6.4.1	162	Photo-z + morphology on bright v.s. faint samples	SDSS Coadd	SDSS, BOSS, VVDS, DEEP2, WiggleZ	ANNZ
6.4.2	168	Photo-z + morphology when star-galaxy separator is bad	SDSS Coadd	SDSS VVDS, DEEP2, WiggleZ	ANNZ
6.4.3	173	Photo-z + morphology with infrared magnitudes	SDSS Coadd, S82-MGC	SDSS, BOSS, VVDS, DEEP2, WiggleZ	ANNZ
6.5.1	179	Photo-z + morphology on the CS82 (weighted sample)	SDSS Coadd, CS82	SDSS, BOSS, VVDS, DEEP2, WiggleZ	ANNZ2
6.5.2	181	Photo-z + morphology with limited no. of filters	SDSS Coadd, CS82	SDSS, BOSS, VVDS, DEEP2, WiggleZ	ANNZ2
6.5.3	184	Photo-z + morphology when photometry is bad	SDSS, CS82	SDSS, BOSS, VVDS, DEEP2, WiggleZ	ANNZ2
7.2	192	$p(z)$, $n(z)$ + morphology	SDSS Coadd, CS82	SDSS, BOSS, VVDS, DEEP2, WiggleZ	ANNZ2, BPZ
7.3	203	Photo-z and morpho-z catalogue for CS82	SDSS Coadd, CS82	SDSS, BOSS, VVDS, DEEP2, WiggleZ	ANNZ2, BPZ
7.4	210	Photo-z, $p(z)$ and $n(z)$ on LSST DC1	Buzzard	Buzzard	ANNZ2, DELIGHT
8.2	228	Photo-z on PAU simulated data	PAU Mock, MICE2	PAU Mock, MICE2	ANNZ, BPZ, ANNZ2, DELIGHT
8.3	239	Photo-z on PAU real data	PAU, COSMOS	zCOSMOS	ANNZ, BPZ, ANNZ2, DELIGHT

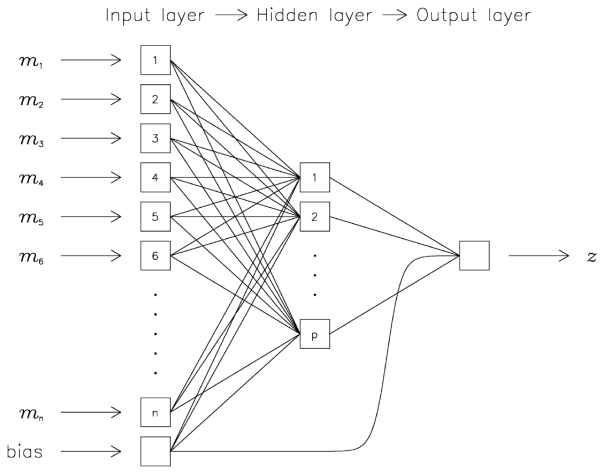

Figure 15: An original figure from Collister and Lahav (2004) showing the structure of an ANN.

5.2.1 *Artificial Neural Network*

ANNz uses ANNs to produce photo-z's. An artificial neural network (ANN), as its name suggests, mimics the network of neurons in the human brain: it is made up of interconnected nodes arranged in several layers. The type of ANN used by ANNz is known as the *multi-layer perceptron* (MLP), which is also the most commonly used. A sufficiently complex MLP is capable of approximating to arbitrary accuracy any continuous functional mapping (Hoecker et al. 2007).

The structure of a MLP used by ANNz is visualised in Figure 15. In a typical setup, the input nodes, which are usually the broadband magnitudes of galaxies, are processed in the inner layers, where a highly non-linear and complex relation is established using a training set of galaxies, and finally the photo-z is estimated in the single output node. The interconnecting layers are called *hidden layers*, and there is complete freedom by the user to set the number of hidden layers and the number of nodes for each hidden layer used. Resources suggest that the rule of thumb is to have the size of hidden layers be between the size of the input and the output layers (Blum 1992) and that it should never be more than twice as large as the input layer (Berry and Linoff 2004).

Each connection between nodes i and j within the network carries a weight w_{ij}, and each node carries a value u_j which is calculated via the summation of weights and activation functions $g_i(u_i)$ of the value u_i of the previous node connected to it, mathematically,

$$u_j = \sum_i w_{ij} g_i(u_i). \tag{77}$$

Therefore in the input layer the broadband magnitudes are the u_i to be acted on by the activation functions of nodes in the first hidden layer, this value is sent to the second hidden layer and the process is repeated, and the photo-z is produced as a result of a system of node values, weights and activation functions. The activation function is usually a sigmoid $\left(\frac{1}{1+e^{-u}}\right)$ or a tanh function, making the process highly non-linear. A bias node as illustrated in Figure 15 allows for additive constants in these functions.

Using the training set of galaxies, the network will constantly adjust the weights such that a cost function E is minimised,

$$E = \sum_i \left[z_{\text{phot},i}(w, m) - z_{\text{spec},i} \right]^2, \tag{78}$$

where z_{spec} and z_{phot} are the spectroscopic and photometric redshifts respectively, with m representing the broadband magnitudes. In the case of ANNZ, the weights are adjusted using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm (Fletcher 1987), a quasi-Newton iterative method best for solving unconstrained non-linear optimisation problems. The weights are updated recursively until the cost function stops improving over a certain threshold value, or when the number of cycles indicated has been reached, thus producing the best photo-z available.

ANNs prevent overtraining by using regularisation on the cost function. A regulator term proportional to the norm of the weights $\alpha|w|^2$ is added to the right hand side of Equation 78, essentially preventing overly large weights which may reduce the complexity of the network. Overtraining is also prevented by separately evaluating the cost function on a validation set after each training iteration, and the training ends when both cost functions of the training and validation set are at their minimum.

5.2.2 *Algorithm*

ANNZ is constructed via the C++ programming language, with its ANN coded from scratch. The code first requires the user to set an architecture specifying the number of inputs, number of nodes in hidden layers and number of outputs. Then a training and validation set are required for the training process, specifying the input and output variables. Multiple trainings can and should be done for each estimation, and each training requires a different random number as input, where different random numbers set different random starting weights. The weights optimised in each run are saved, and the code can be used to test on a testing set to evaluate the photo-z's. A committee of networks (i.e. different saved weights) are recommended to be used for the testing process, where the final photo-z evaluated will be averaged across the results of each training run.

ANNZ derives the uncertainty of its photo-z using two components. The first is the photometric errors from the inputs σ_m (i.e. the magnitude errors) through the linear propagation of errors, and the second is the network variance σ_n, the standard error calculated from the difference in photo-z value estimated by the different committees of the neural networks used. These errors are then combined to produce the photo-z error σ_z,

$$\sigma_z^2 = \sum_i \left(\frac{\partial z}{\partial m_i} \right)^2 \sigma_{m_i}^2 + \sigma_n^2. \tag{79}$$

Therefore when low quality photometry is used, the photometric noise would dominate the error.

5.2.3 *Optimisation*

Most of the earlier photo-z work in this thesis were done by ANNZ, therefore most of the optimisation work on the hyperparameters were done using the SDSS LRG Sample, in an attempt to reproduce the results of the MegaZ-LRG Photo-z Catalogue (Collister et al. 2007). Throughout this work, 4 committees of networks (i.e. 4 training rounds each with random seeds 1, 2 , 3 and 4) were used to obtain the photo-z's. Two different sets of architectures have been used: for Chapter 6 and Chapter 7 the architecture is set to N:2N:2N:1 (N inputs, 2 hidden layers with 2N nodes each, and 1 output) as it was the setting used by Collister and Lahav (2004) and Abdalla et al. (2011). However later in Chapter 8 we switched to N:$\frac{1}{3}(2N+1)$:$\frac{1}{3}(N+1)$:1 recommended by Blum (1992) upon realising that it produced better results than the former. We have tested that the number of hidden layers are optimal and increasing it would only affect the root-mean-square error by about 10^{-4}. For each run we set a maximum of 9999 iterations (to update the weights in the cost function), but the run would usually end well before that when no further improvement in the cost function was found.

Here we note that in all studies in this work, spectroscopic samples are divided equally into 3 sets for training, validating and testing respectively[1]. A common misconception among astronomers is that the size ratio of the training to testing set should be high (e.g. at least 1:10) as this realistically depicts the situation of current surveys, where the spectroscopic training sample is small and the target sample is large. However, it does not work this way: the results of machine learning codes are highly dependent on the absolute sizes of the training and validating sets than the ratio between training and testing. In fact we set the training-validating-testing ratio to 1:1:1 since it would keep the number of training objects as high as possible, much higher than the training size threshold of about 2000 objects suggested by Collister and Lahav (2004) and Bonfield et al. (2010). Large training sets are needed in our work especially for cases where the number of inputs are high (e.g. up to 10 in Chapter 7 and up to 46 in Chapter 8), and smaller training sets would cause biases in the results of for high-weighted objects in weighted samples (Section 6.5). This setting also maximises the testing set such that the metrics evaluated on it are less affected by sample noise.

ANNZ is used mainly in the initial phase of this study (Sections 6.2.1 through 6.4). ANNZ is replaced by ANNZ2 in later sections due to its inability to produce photo-z PDFs and deal with weighted samples. ANNZ is also used in Chapter 8 in comparison with other codes.

5.3 ANNZ2

Despite sharing similar names, ANNZ2 (Sadeh et al. 2016) is independent of ANNZ, and they differ in programming language and functionality. ANNZ2 incorporates the Toolkit for Multivariate Data Analysis (TMVA, Hoecker et al. 2007) with ROOT (Brun and Rademakers 1997) in its code, which is a powerful package that contains a library of machine learning algorithms. Therefore ANNZ2 can run multiple machine learning algorithms for training and outputs photo-z's based on a weighted average of their performances. ANNZ2 has been widely

1 As there are many confusing definitions of validating and testing sets, we remind the reader that in this work the *validating* set is used concurrently with the *training* set to prevent overtraining, while the *testing* set is used to evaluate the performance of the photo-z. See Section 3.2.3 for the full discussion on this topic.

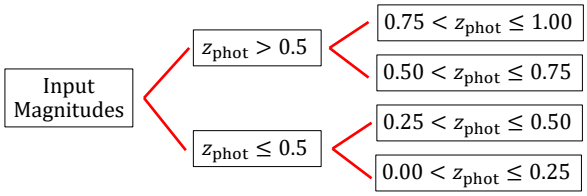

Figure 16: An illustration of a simple decision tree algorithm for photo-z's.

used in survey applications (Sánchez et al. 2014; Bonnett 2015; Jouvel et al. 2017) due to its high customisability and its ability to produce PDFs. ANNZ2 is publicly downloadable from `https://github.com/IftachSadeh/ANNZ`.

Since ANNZ2 is able to use several machine learning algorithms from TMVA in the training process, in this work we have run ANNZ2 with a mixture of ANNs, BDTs and KNNs (and also independently). As ANN has been discussed earlier in Section 5.2.1, in the following sections we describe the machine learning methods BDT and KNN.

5.3.1 *Boosted Decision Tree*

A boosted decision tree (BDT) is a type of decision tree learning algorithm. The term 'tree' here resembles an analogy of a tree where the 'roots' (inputs) are observed, analysed in 'branches' (internal nodes), and concluded by branching them to the 'leaves' (target values), which are the different values the photo-z can assume. A BDT is a binary classifier, it makes a repeated signal-background decision taken on a selected variable until a stop criterion is fulfilled, as illustrated in Figure 16. Since photo-z's are continuous values, decision trees are turned into regression trees by using fine photo-z bins, best chosen to be at least two times smaller than the expected resolution (Gerdes et al. 2010).

The decision tree training process is as follows. First a single decision tree is constructed with the root node consisting of all training objects with their respective input variables. A splitting criteria is chosen such that a single variable cut provides the best separation between signal and background. This is done by first assigning a weight w_i for i objects (initially all set to $w_i = 1$), and scanning over all variables to select a variable cut which maximises the increase in the Gini index G (Breiman et al. 1984) between the parent node and the sum of the Gini indices of its branches, with G defined as

$$G = \left(\sum_{i=1}^{n} w_i \right) p(1-p), \tag{80}$$

where n is the number of objects in that node, while p is the purity of the branch,

$$p = \frac{\sum w_s}{\sum w_s + \sum w_b}, \tag{81}$$

where w_s and w_b are the weights (or number) of the objects in both branches, labelled as the signal and background branches respectively. Similar to ANNs, the weights are updated by minimising the cost function.

In can be deduced from equation Equation 80 that the best split is such that the purity is fully mixed ($p = 0.5$). As the selected variable cut requires the code to scan through all variables, the number of cuts n_c across the grid of variables can be set to prevent long computing time (default value in ANNZ2 is set to be $n_c = 20$). Once a variable cut is selected, the subset will continue to be split with the process repeated, until a node contains the minimum number of objects in the node (or minimum node size) required, and this node is known as the leaf. The default minimum node size in ANNZ2 is set to be 0.02% for the training sample. The cut sequences at each node are remembered and tested on the testing set.

An individual decision tree is a weak classifier (correlates poorly between input and results) and it is very sensitive to statistical fluctuations in the training sample, therefore a small change in the training sample selection may cause a big change in the results. This is solved by using a forest of trees, the overall results will then be based on the weighted performance of regression done by each tree in the forest. All trees are trained using the same training set, however the second and subsequent trees are generated by *boosting* or *bagging*, which increases the statistical stability of the classifier and improves the separation performance over the first tree.

There are many types of boosting algorithms in TMVA, however we will only explain the two types used in this work: adaptive boosting (AdaBoost, Freund and Schapire 1997) and bootstrap aggregating (bagging, Breiman 1996). AdaBoost is done by giving misclassified events a higher weight in the subsequent tree, the weights of misclassified objects are boosted by a factor of $\alpha = \frac{1-R_m}{R_m}$, where $R_m = 1 - \max(p, 1-p)$ is the misclassification rate of a given tree, and the total weights are renormalised again. This makes AdaBoost sensitive to noisy data and outliers, an advantage which ANNs do not have. As our case of photo-z is a regression problem, the AdaBoost.R2 algorithm (Drucker 1997) is used instead. This algorithm works in a similar way as AdaBoost, however the boosting factor is defined using different forms of loss functions. The mathematics of AdaBoost.R2 are too complicated to be discussed here, the reader can refer to Drucker (1997) for its full details.

Bagging, on the other hand, builds multiple decision trees by repeatedly resampling training data with replacement. Bagging is technically not a boosting technique, however it combines several resampled trees into a collection which results in an averaged classifier that has less statistical fluctuation than using one decision tree. By default, TMVA resamples 60% of the training sample and randomly assigns Poisson weights to each event of the parent sample. Both AdaBoost and bagging produce multiple decision trees between 20 to 100, which produces a forest of trees. The final photo-z for an object is calculated by the weighted average leaf of each tree.

BDTs have many advantages: they are simple to understand and interpret, they make no assumption on the training data, they run pretty quickly, they perform well with large datasets, and are more sensitive to correcting outliers. However the disadvantage is that at times its result tend not to be as accurate as those produced by ANN due to a lower complexity in its architecture.

5.3.2 *K-Nearest Neighbours*

The K-nearest neighbours (KNN) regression algorithm is one of the simplest machine learning algorithms available. It uses the properties of a fixed number (K) of nearest neighbours in the input parameter space to produce the output parameter. In the case of photo-z's, the input parameter space of the training galaxies is the magnitude space (e.g. SDSS *ugriz*). Considering coordinates in the magnitude space, to find the photo-z of a galaxy at coordinate y_i in the testing set, the algorithm searches for K galaxies at coordinate x_i in the training set that are closest to the the galaxy at y_i with respect to a Euclidean distance R_{E}, defined as

$$R_{\mathrm{E}} = \sqrt{\sum_{i=1}^{b} |x_i - y_i|^2}, \tag{82}$$

where b is the number of bands used. The K objects with the smallest R_{E} to the test galaxy are its K-nearest neighbours. As simple as it sounds, the photo-z of the test object is equal to the average of the spec-z of its K-nearest neighbours in the magnitude space.

If the input variables are in different units (e.g. when morphological parameters are used in training in addition to magnitudes), the KNN tends to be more sensitive to the variables with a wider distribution as dictated by R_{E}, therefore weights need to be assigned to normalise each variables so that they have the same scale. A weight w_i can be applied to Equation 82 to give

$$R_{\mathrm{E}} = \sqrt{\sum_{i=1}^{b} \frac{1}{w_i} |x_i - y_i|^2}. \tag{83}$$

A scale fraction can be set such that the weight of each variable is scaled using only a certain fraction of objects from the median, this is to prevent the weights being scaled with the presence of outliers. The default scale fraction used in TMVA is 80%.

Just like BDTs, KNNs are also very sensitive to the training data and the value of K, since a slight change of K may give very different results. The high variability can be mitigated by adding a weight function with a polynomial kernel W such that

$$W\left(\frac{R_{\mathrm{E},i}}{R_{\mathrm{E},K}}\right) = \begin{cases} \left(1 - \left|\frac{R_{\mathrm{E},i}}{R_{\mathrm{E},K}}\right|^3\right)^3 & \text{if } R_{\mathrm{E},i} < R_{\mathrm{E},K}, \\ 0 & \text{otherwise.} \end{cases} \tag{84}$$

where $R_{\mathrm{E},K}$ is the distance to the Kth neighbour. The weight function gives a higher weight to closer neighbours, giving more accurate results.

The value of K is arbitrary, but it has to be small enough to capture local behaviour, yet large enough to prevent the inclusion of statistical fluctuation and noise from the sample. Typical values of K are between 10 and 100 for a sample of thousands of objects, the larger the training set the better, as it allows the algorithm to probe small-scale features. The KNN method is not appropriate for problems when input parameters exceed 10, as the Euclidean distance of vectors in such a high dimensional space may turn out to be almost equidistant during the search query.

Dimension reduction can be performed by conducting a PCA of the variables, replacing them with their eigenvectors.

5.3.3 *Algorithm*

ANNZ2 is a python code which has 3 different modes: single regression, randomised regression and classification. Single regression mode is essentially just like ANNZ where it outputs a single point estimate photo-z for each galaxy using a single machine learning method, while classification is suitable for star-galaxy separation or morphological classifications. Randomised regression is the most suitable mode for this work, as it can produce photo-z's using multiple runs with different machine learning algorithms or settings, and also outputs photo-z PDFs. First it processes the input training, validation and testing sets and saves them in formats to be processed by TMVA. Next it trains the data, with the user required to set the type of machine learning algorithms and their respective hyperparameter settings, and whether weights are used in the training process. Finally it evaluates the photo-z on the testing set and saves the best point estimates and photo-z PDFs in a single file.

ANNZ2 produces 4 point estimate photo-z's for each run: $z_{\rm best}$ is the photo-z produced using only the best run; $z_{\rm mlm}$ is the mean value of the PDF produced via an unweighted average of all runs; while $z_{\rm pdf}$ and $z_{\rm peak}$ are the mean and peak value of the PDF produced via a weighted average of all runs.

In deriving the photo-z errors, ANNZ2 does not use the propagation of magnitude errors, instead it derives uncertainties using the KNN method. It first finds K nearest neighbours (in magnitude space) of the galaxy in the training set, calculates the photo-z biases $\Delta z_K = z_{\rm phot} - z_{{\rm spec},K}$ between these objects, and the photo-z uncertainty of the target object is taken to be the 68th percentile width of the distribution of the bias. This asserts that objects with similar photometric properties should have similar uncertainties (Sadeh et al. 2016).

For photo-z PDFs, ANNZ2 derives them by first producing a single-value photo-z solution for each run, it is then combined with its corresponding photo-z uncertainty (using the KNN method) to produce a Gaussian PDF. The final PDF is obtained by weighting the different PDFs produced by each run, this has the advantage of exposing configurations which perform badly due to poor choice of algorithm parameters or statistical fluctuation. The different runs are weighted by comparing the cumulative distribution function (CDF) at the spec-z, namely

$$C(z_{\rm spec}) = \int_0^{z_{\rm spec}} p(z)\, dz \tag{85}$$

where the $p(z)$ is the normalized photo-z PDF. Essentially $C(z_{\rm spec})$ is the percentile of the spec-z in the $p(z)$, and we expect the distribution of $C(z_{\rm spec})$ to be as flat as possible since we expect $z_{\rm spec}$ to be a random variable distributed across the PDF. Therefore the machine learning estimator with a flatter distribution of $C(z_{\rm spec})$ would have a higher weight and vice versa.

ANNZ2 also addresses the uncertainty of an unrepresentative training set via reweighting, as discussed in Section 3.3.4. Weights for individual galaxies can be input manually or calculated using the code itself, and the cost function will be optimised with respect to the weights. The

code also allows an output of a flag to indicate that an object has unreliable photo-z if it falls within an incomplete region in the training sample.

5.3.4 *Optimisation*

In this work, ANNZ2 version 2.0.4 was used, its hyperparameters were optimised mainly using the CS82 Sample (Section 6.5) and the PAU Mock Galaxy Sample (Section 8.2.1). $z_{\rm pdf}$ is chosen to represent the point estimate photo-z of ANNZ2 as it is found to be more accurate than $z_{\rm peak}$. $z_{\rm peak}$ is however the estimate used to determine the odds value Θ of an object (see Equation 71). For the photo-z error estimation we use the default value $K = 100$, it is selected so that it is large enough for the uncertainty not to be limited by shot noise, and small enough to ensure the estimate remains local in the input parameter space. As for samples with weights, we do not use the inbuilt code to generate weights but instead use an external programme to do so and manually input them into the sample.

The ANN settings and architecture used for ANNZ2 are similar to those used in ANNZ, however ANNZ2 has a higher customisability than ANNZ. After optimising for best results, we set the number of maximum cycles to 6000, the convergence of improving the cost function as 10^{-30}, neuron type as a tanh function, training method to BFGS with a reset step of 250, overtraining tested on every 25th run, convergence tested on every 25th iteration, remove the use of the regulator, and normalising all variables before training and testing.

For BDTs we chose to only use bagging but not AdaBoost, since we found that it generally produces at least 14% less scatter when compared to AdaBoost. We have set the minimum node size to 2%, number of variable cuts $n_c = 50, 100$, number of trees between 5 and 20, and normalisation and PCA are done on all variables prior to training. Finally for KNNs we used a polynomial kernel, set K between 5 to 25, and a scale fraction of 1.0.

ANNZ2 is used in every section beginning with Section 6.5, mostly in situations when $p(z)$ and reweighting are required. In this study, 5 runs are conducted for each training round if a single type machine learning method is used, with each run using different hyperparameter settings as described above. If a mixture of machine learning methods are used, 2 ANN and BDT runs each and 1 KNN run will be conducted. ANNZ2 results from sections prior to Chapter 8 were produced with only ANNs, while results after that include a mixture of them.

5.4 BPZ

Bayesian Photometric Redshift (BPZ, Benítez 2000) is a widely used template-based photo-z code by the photo-z community. We have briefly discussed the various functionalities of BPZ earlier in Chapter 3, particularly its use of priors in constructing photo-z PDFs and its use of an odds parameter Θ as a photo-z quality cut (see Section 3.3.2). Therefore we will not repeat the discussion of these issues in this section. BPZ is used by many including the works of Hildebrandt et al. (2008), Hildebrandt et al. (2010), Abdalla et al. (2011), Martí et al. (2014), Bundy et al. (2015) and Cavuoti et al. (2017a). BPZ is publicly downloadable from `http://www.stsci.edu/~dcoe/BPZ/`.

5.4.1 *Algorithm*

BPZ compares the expected magnitudes calculated for a set of SED templates to the actual observed magnitudes m, in order to calculate the *likelihood of observing magnitudes* $p(m|z,t)$ at redshift z and template type t. Assuming that the SED templates t are exhaustive, the *redshift posterior* $p(z|m)$ is determined by marginalising over all templates with a simple sum:

$$p(z|m) \propto \sum_t p(m|z,t)p(z,t), \tag{86}$$

where $p(z,t) = p(t)p(z|t)$ is the Bayesian prior which describes the redshift and spectral type distribution. The redshift likelihood $p(m|z,t)$ can be obtained via fitting the observed fluxes to the template fluxes,

$$p(m|z,t) \propto \exp\left[-\frac{1}{2}\sum_i^n \frac{(F_i - aF_t)^2}{\sigma_{F_i}^2}\right], \tag{87}$$

where F and F_t are the observed and template fluxes respectively, summed over all galaxies. σ_F is the flux error, while a is a normalising constant to account for the overall brightness of the galaxy.

The prior $p(z,t)$ as discussed in Section 3.3.2 is able to reduce the outlier fraction rate of the photo-z since it tells us the expected distribution of z and t, thus suppressing likelihoods which sometimes points towards absurdly high redshifts. $p(t)$ sums up to one, in a flat prior cases we assume and that galaxies belonging to each template are equally probable, therefore having equal values of $p(t)$ for every template. $p(z|t)$ on the other hand is the expected redshift distribution for each template, which in a flat prior case all templates are assumed to have $p(z|t)$ equal to the redshift distribution in the training set.

However, we know that flat priors are not realistic since we expect starburst and AGN templates to be less probable, therefore the prior $p(z,t)$ can be constructed based on spectroscopic observations, e.g. using data from HDF-North (HDF-N, Williams et al. 1996) or COSMOS (Laigle et al. 2016). For example, the spectral fractions at Hubble magnitude $F814W = 20$ have a distribution of 35% ellipticals, 50% spirals and 15% irregulars, respectively, and using a simple linear magnitude dependence on redshift, this can be used to constrain the final $p(z)$ to remove unrealistic outliers at different redshifts.

BPZ also allows interpolation between templates, thus making each template not exclusive. This can be done by setting a number to interpolate between an ordered set of templates (ellipticals, spirals and others). Benítez (2000) has shown that significant improvement in photo-z can be observed when the number of interpolations between templates is set to 2.

5.4.2 *Optimisation*

In this work, BPZ version 1.99.3 was used, the code was optimised using the PAU-MICE2 Simulated Sample. We assume the same functional form for the Bayesian priors used by Benítez (2000) (the HDF-N priors), but the prior for COSMOS (Laigle et al. 2016) was preferred for PAU

samples. In both cases, the i-band for each sample is used as the reference band to process the priors. For the photo-z point estimates, z_b (the peak of the PDF) was used.

For samples using galaxy fluxes rather than magnitudes (e.g. samples in Chapter 8), BPZ requires that the fluxes to be levelled to $F = 10^{-0.4m}$ and the non-detection fluxes set to 0. We also set the template interpolation parameter to 2. Other than those stated, the default settings were used. The zero-point offsets for each band produced by BPZ after each run are input back into the code, and the code is run once more for better results.

BPZ is used only in Section 7.2 and Chapter 8, primarily used just for comparison with ANNZ, ANNZ2 and DELIGHT.

5.5 DELIGHT

DELIGHT (Leistedt and Hogg 2017) is a hybrid photo-z algorithm between template-based and machine learning. Generally, machine learning photo-z's are produced relying on representative training data with the same band passes, while template based methods rely on a complete library of templates based on physical models constructed. DELIGHT however, is constructed to combine the advantages and eliminate the disadvantages of both template-based and machine learning algorithms: it constructs a large collection of latent SED templates (or physical flux-redshift models) from training data, with a template SED library as a guide to the learning of the model. This conceptually novel approach is done by using GPs operating in flux-redshift space, which will be described in Section 5.5.1.

DELIGHT is a relatively new code, it has not been applied in any published work yet, but will feature in the LSST Photo-z Data Challenge 1 (Schmidt et al., in prep.). DELIGHT is publicly downloadable from `https://github.com/ixkael/Delight`.

5.5.1 *Gaussian Process*

A Gaussian process (GP) is considered a supervised learning method, it finds a distribution over the possible functions $f(x)$ that are consistent with the observed data x. Consider Figure 17: suppose we have a set of observed variables $y = f(x)$ to which we would like to fit a function. Instead of fitting a straight line or a polynomial, we fit it using a GP, $f \sim \mathcal{GP}(\mu, k)$, which assumes that the probability of all $f(x)$ is jointly Gaussian and representable by a mean function $\mu(x)$ and a covariance matrix $\Sigma(x) = k(x_i, x_j)$, where $k(x_i, x_j)$ is the kernel function, which is a function that relates one variable x_i to another x_j. In the simplest case, it is usually assumed that the mean function is zero, and the kernel function takes the form of a squared exponential,

$$k(x_i, x_j) = \sigma_f^2 \exp\left[\frac{-(x_i - x_j)^2}{2l^2}\right], \tag{88}$$

where σ_f^2 is the maximum allowable covariance between data, and l a length parameter of normalisation. Leaving $\mu = 0$ and k the squared exponential in this case essentially means that the GP will try to find a marginalisation of all possible functions, however μ and k can be modified

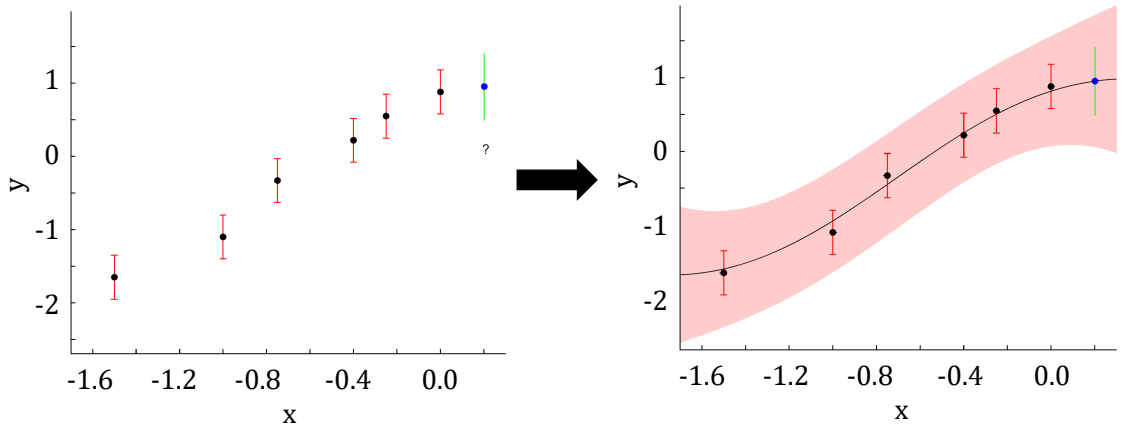

Figure 17: Illustration of a Gaussian process. On the left shows given data points (red error bar), with a single data to be predicted (green error bar). The Gaussian process would train on the given data points to provide a best fit function (black line) as shown on the right, it would also provide a Gaussian confidence interval (pink area) for the function. Figure obtained and modified from Ebden (2015).

if an underlying model of the data that we want to fit is known, as it will be shown later in Section 5.5.2.

Essentially the covariance function is defined such that variables x_i and x_j that are near each other should have $f(x_i)$ and $f(x_j)$ near each other too, which in fact defines a smooth function to be predicted. Now assume that we have a set of training data $\{x_i, f(x_i)\}$ and would like to find the prediction $\{x_*, f_*(x_*)\}$. GP assumes that f and f_* are jointly Gaussian, therefore

$$\begin{pmatrix} f(x) \\ f_*(x) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu \\ \mu_* \end{pmatrix}, \begin{pmatrix} \Sigma & \Sigma_* \\ \Sigma_*^T & \Sigma_{**} \end{pmatrix} \right), \tag{89}$$

where $\Sigma = k(x_i, x_j)$ is the covariance between the training data, $\Sigma_* = k(x_*, x_i)$ the covariance between training and the predicted data (superscript T denoting the transposition), and $\Sigma_{**} = k(x_*, x_*)$ the covariance between testing data.

Skipping the complex mathematical proof, it follows that the posterior $p(f_*|x_*, x_i, y_i)$ follows a Gaussian distribution too, and can be expressed as

$$p(f_*|x_*, x_i, y_i) \sim \mathcal{N}(\bar{f}_*, \sigma_{f_*}^2) = \mathcal{N}(\mu_* + \Sigma_* \Sigma^{-1}(f - \mu), \Sigma_{**} - \Sigma_* \Sigma^{-1} \Sigma_*^T), \tag{90}$$

where $\bar{f}_*(x_*) = \mu_* + \Sigma_* \Sigma^{-1}(f - \mu)$ is the best estimate of f_* given x_*, while $\sigma_{f_*}^2 = \Sigma_{**} - \Sigma_* \Sigma^{-1} \Sigma_*^T$ is its variance. The predicted blue dot and green error bars in Figure 17 representing $f_*(x_*) \pm \sigma_{f_*}$ is calculated this way, with the black line the smooth predicted function $\bar{f}_*(x)$ across all points. The pink shaded area shows the 95% confidence interval ($\pm 1.96\sigma_{f_*}$) of $\bar{f}_*$.

In the context of DELIGHT, the GP is used to calculate the predicted fluxes in any required band of a training object, but at a different redshift, which we would explain in more detail below.

5.5.2 *Algorithm*

DELIGHT produces the posterior photo-z distribution of a target galaxy via a pairwise comparison with training galaxies. The posterior distribution $p(z|F)$ of an object in the testing set with redshift z given the set of photometric fluxes F has a similar expression to Equation 86, however instead of referring to SED templates, t_i is constructed from the ith training object (at redshift z_i and flux F_i), therefore

$$p(z|F) \approx \sum_i p(F|z,t_i)\, p(z|t_i)p(t_i), \tag{91}$$

where $p(z|t_i)p(t_i)$ is the prior and $p(F|z,t_i)$ is the probability of the flux F at redshift z.

For each training-target pair, $p(F|z,t_i)$ is evaluated as follows:

$$\begin{aligned} p(F|z,t_i) &= p(F|z,z_i,F_i) \\ &= \int p(F|F_{\rm pr})\, p(F_{\rm pr}|z,z_i,F_i)\, dF_{\rm pr}, \end{aligned} \tag{92}$$

where $p(F|F_{\rm pr})$ is the likelihood function which compares the flux F from the target object with the predicted flux $F_{\rm pr}$ from training; while $p(F_{\rm pr}|z,z_i,F_i)$ is the prediction of flux $F_{\rm pr}$ at the redshift of the target object z with respect to the training object with redshift z_i and flux F_i. The likelihood function can be derived using a Gaussian distribution,

$$p(F|z,z_i,F_i) = \int \mathcal{N}(F - \ell F_{\rm pr}, \Sigma_F + \ell^2 \Sigma_{F_{\rm pr}}) p(\ell)\, d\ell, \tag{93}$$

where $p(\ell)$ is the prior with ℓ a factor to scale the predicted flux and its uncertainties. $p(\ell)$ is also modelled as a Gaussian with $p(\ell) = \mathcal{N}(1;\sigma_\ell^2)$.

The predicted flux $F_{\rm pr}$ for the training galaxy at target redshift z in Equation 93 is modelled via a GP, $F_{\rm pr} \sim \mathcal{GP}\,(\mu,k)$ with mean function μ and kernel k, both imposed to capture expected correlations resulting from the known underlying physics (i.e., fluxes resulting from observing SEDs through filter response, and the SEDs being redshifted). This means that we are trying to find $p(F_{\rm pr}|z,z_i,F_i)$, the probability of the predicted flux $F_{\rm pr}$, given a certain redshift z and a set of training data $\{z_i,F_i\}$. In simple words, we intend to compute the probability that the target galaxy has the same SED as the training galaxy at a different redshift, and therefore the GP is used to create latent SEDs in flux-redshift space from our training data, and predict fluxes to be compared to the fluxes in the testing set. Therefore the μ and k are no longer zero or a merely squared exponential as they have to be set in a way such that the predicted fluxes are results from observing a redshifted SED through a given filter response.

With different bands b, different underlying SED templates t and different flux normalisations between SED templates ℓ, the multivariate GP can be written as

$$F_{{\rm pr},b}(z,t,\ell) \sim \mathcal{GP}\left(\mu(b,z,t), k(b_i,b_j,z_i,z_j,\ell_i,\ell_j)\right). \tag{94}$$

The mathematical equations and derivations of μ and k are way too complex to be written down in this work, we refer the reader to Leistedt and Hogg (2017) for the full equations and derivations. Qualitatively, the mean function μ is now a function of redshift, a linear mixture

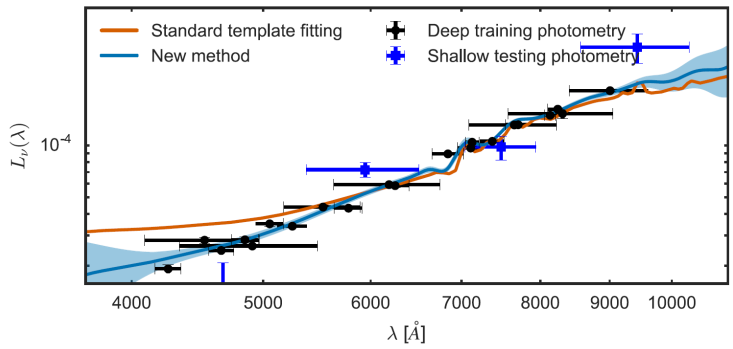

Figure 18: An original figure from Leistedt and Hogg (2017) illustrating how the Gaussian process 'corrects' the SED templates using training data to get better photo-z's. Note the use of different bands in training and testing.

of SED template sets and the corresponding filter responses, allowing it to make predictions to follow the physics of known SEDs; the kernel k on the other hand is a function of the redshift, filter responses of each band and flux normalisations across different templates, which makes 'smooth' predictions of the fluxes values between band filters and redshift. Within the kernel itself is in fact another kernel which controls the variance and smoothness of the continuum and emission line residuals in addition to the given SED templates.

The process of the photo-z estimation of DELIGHT can be summarised as follows. Firstly, all filter responses R_λ for each band are modelled as a linear combination of n Gaussians ($R_\lambda = \sum_i^n w_i \mathcal{N}(\lambda - \mu_i, \sigma_i^2)$ with w as a weight), where n can be set by the user. Next the library of SED templates are processed such that for each template, the predicted flux value for each band at each redshift (in stepsizes set by the user) is recorded. The filters and SED templates are processed as such to ease the computation of the GP later. DELIGHT then proceeds to learn from the training data, creating a GP for each training object. The GP for each training galaxy is saved, and finally applied to the testing data, predicting fluxes of test galaxies for each training-testing pair (Equation 90), finding the probability that the test galaxy has the same SED as the training galaxy but at a different redshift (Equation 92), and at last the photo-z and its PDF for each galaxy is produced (Equation 91), which is a weighted solution based on the best fit of each training-testing pair. At the same time, DELIGHT also produces a separate photo-z for each galaxy based on a simple template fitting method (which will be referenced as the DELIGHT template code); both photo-z's produced can then be compared and analysed.

Figure 18 illustrates how the GP improves on existing template fitting methods. First DELIGHT is trained on deep Subaru and Hubble Space Telescope (HST) photometry from COSMOS (black dots), creating a Gaussian process (blue line with shaded confidence interval) based on a standard template (orange line), which is then tested on shallow SDSS photometry (blue dots) to find the best fit photo-z. Note the discrepancy between the orange and blue line, showing the difference between template fitting and empirical latent SEDs produced by DELIGHT.

As illustrated in Figure 18, one of the greatest advantages of DELIGHT is that it neither needs representative training data in the same photometric bands, nor does it need detailed galaxy SED models to work. It just requires accurate spectroscopic redshifts and and high quality training fluxes to work. As such, given a few photometric bands, DELIGHT is able to predict missing

bands or fluxes, and this function is utilised in Section 8.3.1 to calibrate the fluxes between two surveys.

5.5.3 *Optimisation*

Throughout this work, we use z_{map} (the *maximum a posteriori* of the PDF) to represent the best point estimate photo-z produced by DELIGHT. Default settings of DELIGHT were used, with the exception that the PDF bins were set to be linear instead of logarithmic, with photo-z PDF binning sizes of $\delta z = 0.01$ for Section 7.4 and $\delta z = 0.005$ for Chapter 8. The number of Gaussians to fit broadband filter curves is set to 20, while set to 7 for narrowband filters, appropriately selected based on the FWHM of the filters.

All magnitudes and magnitude errors have to be converted into fluxes and flux variances to be processed in DELIGHT. However, unlike BPZ, the fluxes F in DELIGHT have different units, they are related to the magnitude m via the equation $F = 10^{-0.4(m-26.4)}$ (i.e. the magnitudes require a zero-point adjustment of 26.4). DELIGHT is very sensitive to the flux variances, therefore in certain cases a small percentage of flux is added in quadrature to the flux variances. When not stated, the flux added is 0.001%.

While DELIGHT is able to run with negative flux values (non-detections), it has to run with a reference band with positive flux values to normalise the fluxes. For Section 7.4 the LSST i-band is used as the reference band, while for Chapter 8 the PAU narrowband $nb625$ is used as the reference band when narrowbands are used, while the broad r-band is used if otherwise.

Priors in DELIGHT can be set manually, in fact it even comes with a code which would optimise the values of $p(t)$ and $p(z|t)$. In contrast to BPZ we choose to set a flat prior, setting $p(t) = \frac{1}{n_t}$ where n_t is the number of templates used, while $p(z|t)$ for each template is set to the peak of the spec-z distribution of the training set. Note that this setting has negligible impact on the GP code as it moves away from the prior in the process.

5.6 SED TEMPLATE SETS

Galaxy SED template spectra are usually derived from local low-redshift galaxies, representing galaxies from different morphological classes (ellipticals, discs, irregulars, starbursts etc). As BPZ and DELIGHT rely on the use of SED templates to produce photo-z's, we have explored and used 6 template sets in this work. In the following sections we briefly describe these template sets, and the list of templates and their sources are summarised in Table 8. The first five template sets are used in Chapter 8 while the Buzzard template set is used in Section 7.4.

5.6.1 *BPz*

The BPZ template set is the default template set provided by BPZ. It was originally made up of 6 SED templates (originally known as the CWW+SB template set), it was later re-calibrated by Coe et al. (2006) and 2 more templates were added to it.

The first 4 templates (E/S0, Sbc, Scd and Im) originated from those introduced by Coleman, Wu and Weedman (CWW, Coleman et al. 1980), these are intrinsic SEDs of 4 different galaxy

Table 8: Summary of the 6 template sets used by BPZ and DELIGHT in this work, listed are the names of template sets, number of templates in the set, and their sources.

Template set	No. of templates	Source
BPZ	8	Coleman et al. (1980), Kinney et al. (1996), Bruzual and Charlot (2003)
CWW-Kinney	10	Coleman et al. (1980), Kinney et al. (1996)
LEPHARE	66	Coleman et al. (1980), Kinney et al. (1996)
COSMOS	31	Bruzual and Charlot (2003), Polletta et al. (2007)
Brown	129	Brown et al. (2014)
Buzzard	150	DeRose et al. (in prep.)

morphological types constructed using population synthesis. Their ultraviolet spectra were constructed using ultraviolet colours of local galaxies which were observed using the Astronomical Netherlands Satellite (ANS, Wesselius et al. 1982).

The next 2 templates (SB2 and SB3) originate from Kinney et al. (1996), these are starburst SEDs based on observations of local galaxies by the International Ultraviolet Explorer (IUE, Macchetto 1976), built using matched aperture photometry from ultraviolet and optical spectra, therefore providing consistent spectral information between 1200 to 10 000 Å. These templates were included to accommodate faint blue galaxies.

The last 2 templates (SSP5Myr and SSP25Myr) added by Coe are simple stellar population (SSP) models from Bruzual and Charlot (2003). These are synthetic templates that were constructed using the isochrone synthesis code GISSEL, based on the library of observed stellar spectra STELIB (Le Borgne et al. 2003). These SSP templates describe spectral evolution of stellar populations at different ages and metallicities across wavelengths ranging between 91 Å and 160 μm with a resolution of 3 Å. These were added to accommodate even bluer galaxies than those represented by the starburst templates.

This template set was first used in Coe et al. (2006) to obtain photo-z's in the Hubble Ultra Deep Field (HUDF). This template set is known to badly represent LRGs (Abdalla et al. 2011).

5.6.2 *CWW-Kinney*

The CWW-Kinney template set is very similar to the BPZ template set. It consists of 10 templates, with the same 4 CWW templates but extrapolated in the ultraviolet and infrared using theoretical SEDs from the GISSEL library (Bruzual and Charlot 2003), and 6 Kinney starburst templates (SB1 to SB6), using the calibrated versions from Calzetti et al. (1994). This template set was first used by Arnouts et al. (1999) to compare the different template sets to obtain photo-z's for the HDF-N, and it was later provided in the template library repository of LEPHARE.

5.6.3 *LePhare*

The LEPHARE template set (also known as the New CWW Extended Templates) was introduced by Ilbert et al. (2006) to obtain photo-z's for the CFHT Legacy Survey (CFHTLS), and is the

default template set for the LEPHARE photo-z code. This template set consists of 66 templates derived from the original 4 CWW templates and 1 Kinney starburst template. In Ilbert et al. (2006) the 4 CWW templates were optimised and rescaled with training data from VVDS, and were interpolated from one another to form the remaining 61 templates in the set. The starburst template however was not optimised in order to retain the representation of emission lines through this template.

5.6.4 *COSMOS*

The COSMOS template set was introduced by Ilbert et al. (2009) to obtain photo-z's for the COSMOS field using 30-band photometry. This template set consists of 31 templates, 3 elliptical and 6 spiral (S0, Sa, Sb, Sc, Sd and Sdm) templates were from Polletta et al. (2007), which were generated using the chemical evolution code GRASIL (Silva et al. 1998). These templates were selected to replace the CWW templates as they were found to provide better joining of ultraviolet and mid-infrared. These templates belong to a library of SEDs used to fit X-ray observed AGNs by the X-ray multi-mirror mission Medium Deep Survey (XMDS, Chiappetti et al. 2005). These 9 templates were further interpolated to form 19 of the 31 templates in this set.

The remaining 12 are SSP templates from Bruzual and Charlot (2003), with ages ranging from 0.03 to 3 Gyrs. On top of everything, extinction laws were incorporated into the templates as well, using the extinction curve of Prevot et al. (1984) for templates redder than SB3, and the curve of Calzetti et al. (2000) for templates bluer than SB3. We note that in this work we utilise this template set without applying any extinction curves to them.

5.6.5 *Brown*

The Brown et al. (2014) template set consists of 129 SEDs derived from real nearby galaxies. These templates have wavelengths covering from ultraviolet to mid-infrared, covering a broad range of galaxy types including ellipticals, spirals, merging galaxies, blue compact dwarfs and luminous infrared galaxies. These templates differ from those mentioned earlier as they are real galaxy templates. They also differ by the use of a larger aperture size to measure and extract the spectra, since the spectra of both Coleman et al. (1980) and Kinney et al. (1996) were of galaxy nuclei, which may not be representative of integrated galaxy spectra. Brown templates are constructed using optical spectroscopy from Moustakas and Kennicutt (2006) and mid-infrared spectroscopy from Spitzer (Werner et al. 2004) and Akari (Murakami et al. 2007), with their spectra verified using matched aperture photometry of 26 bands from various surveys including SDSS, GALEX and 2MASS.

5.6.6 *Buzzard*

The Buzzard template set is used specifically for the Buzzard Simulated Galaxy Catalogue (DeRose et al., in prep.). As described in Section 4.2.1, each simulated galaxy has an SED drawn from an empirical set of real galaxy spectra from the NYU-VAGC. 150 templates were derived from the SEDs of the sample by using a k-means clustering algorithm (in this case $k = 150$).

As the SEDs of the sample are represented by 5 weights, the k-means cluster centres span the space of these coefficients to properly reflect the underlying density in the coefficient space, then these 150 centres were taken as the weights to construct 150 SEDs using KCORRECT (Blanton and Roweis 2007). These templates are rank ordered such that a set of the first N templates will be the N most spanning templates for the given 150 templates.

5.7 PHOTOMETRIC REDSHIFT PERFORMANCE METRICS

Throughout this work we evaluate the photo-z results of different codes using multiple metrics of performance. In this section we define all the metrics used in this work to quantify the performances of photo-z's, for point estimates, PDFs and redshift distributions.

5.7.1 *Redshift Point Estimates*

In this work we use 3 metrics to assess the accuracy of our point estimate photo-z's produced. The first is the *root-mean-square error* $\sigma_{\rm RMS}$, defined as

$$\sigma_{\rm RMS} = \sqrt{\frac{\sum w_i \Delta z_i^2}{\sum w_i}} \,, \tag{95}$$

where w_i is the weight of the object , and $\Delta z_i = \frac{z_{{\rm phot},i} - z_{{\rm spec},i}}{1+z_{{\rm spec},i}}$, is the difference between the photometric and spectroscopic redshift, scaled by $1 + z_{\rm spec}$. The weights are used in weighted samples when the training set is not representative of the testing set (particularly results from Section 6.5 to Section 7.2). The weights are used in training and also for metric evaluation to forecast the photo-z performance of the target set. In an unweighted sample, all weights are simply $w_i = 1$. Note that $\sigma_{\rm RMS}$ is calculated without outliers removed, and thus measures the overall scatter of the sample.

Next is the 68*th percentile error* σ_{68}, it is the half width of the distribution of Δz_i containing 68% of the objects,

$$\sigma_{68} = \frac{Q_{84.1\%}(\Delta z_i) - Q_{15.9\%}(\Delta z_i)}{2}, \tag{96}$$

where Q are the quantiles of the distribution of Δz_i. In the case of a weighted sample, weighted quantiles are used. σ_{68} measures the core width of the photo-z distribution, with reduced sensitivity to outliers.

Lastly we have the *outlier fraction rate* $\eta_{\rm out}$, this is the weighted / unweighted percentage of objects for which

$$|\Delta z_i| \geq 0.15 \,, \tag{97}$$

as introduced by Ilbert et al. (2006). This metric identifies the density of outliers, the percentage of objects having catastrophic photo-z's. The specific threshold value 0.15 is chosen to enforce consistency with previous literature.

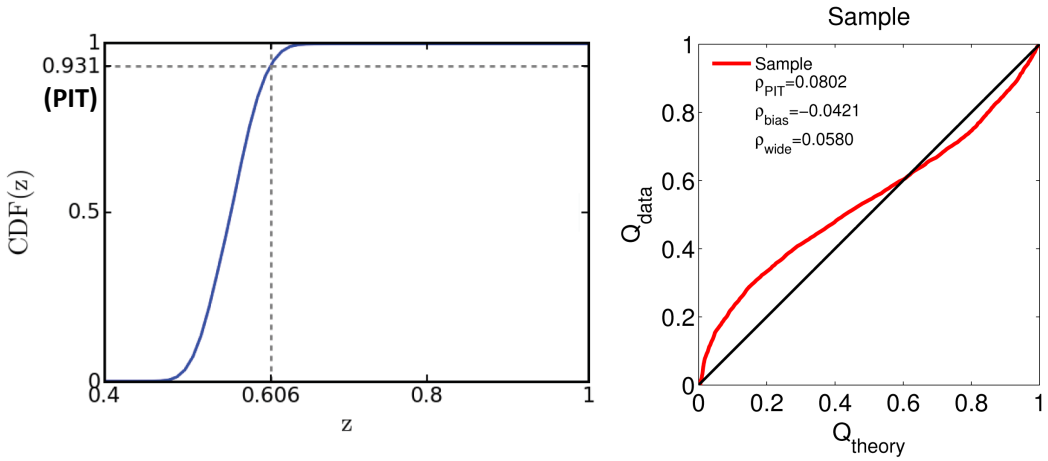

Figure 19: An illustration of a sample QQ plot. First the PIT values of every photo-z CDF is measured (left), a normalized cumulative distribution of these values are tabulated, and the QQ plot (right) is literally the said cumulative distribution when the bars are plotted horizontally. The plot shows a curve with a 'Z' shape, implying that on average the PDFs produced are too wide.

5.7.2 *Redshift Probability Density Functions*

We characterise photo-z PDFs in this work using 2 methods. The first is the *quantile-quantile (QQ) plot*, it is a graphical method for comparing two distributions to analyse the characteristics of the produced photo-z PDFs qualitatively. To construct a QQ plot, we first need to calculate $C(z_{\rm spec})$ as shown in Equation 85, the percentile of the CDF at $z_{\rm spec}$ while asserting the $p(z)$ to have an area of unity. $C(z_{\rm spec})$ is also known as the *probability integral transform* (PIT). The cumulative distribution histogram of these PIT values are tabulated, and the QQ plot is technically such a histogram, except that the x- and y-axes are flipped, and with the histogram bin sizes as small as possible (such that a smooth line can be drawn across the top of each bar).

Figure 19 shows a sample of a QQ plot, with $Q_{\rm data}$ representing the cumulative distribution of the PIT, and $Q_{\rm theory}$ representing the quantiles of the distribution. The overall shape of the $p(z)$ can be evaluated by assuming that for x% of the sample, we expect x% of the $p(z)$ in the data has covered $z_{\rm spec}$ within its PDF. Therefore in an ideal situation, a sample with perfectly constructed $p(z)$ should follow the diagonal black line in Figure 19. If the QQ curve has an 'S' shape, it means that on average the $p(z)$ are too narrow, and a 'Z' shape dictates otherwise. From the QQ plot we can also tell the bias of the photo-z PDFs produced: if the curve is predominantly below the diagonal line, it means that on average most of the PDFs have over-predicted the redshift, and the redshift is under-predicted if otherwise. Since there is no 'true' $p(z)$ to compare the produced $p(z)$ with, the QQ plot gives us an overall view on how well the shapes of the $p(z)$ produced in a collective manner.

Based on the shapes of the QQ plots described above, we introduce a few quantitative measures of the QQ plot:

1. The *root-mean-square of the distribution of the PIT value* $\rho_{\rm PIT}$,

$$\rho_{\rm PIT} = \sqrt{\frac{1}{n}\sum(Q_{\rm theory} - Q_{\rm data})^2}, \tag{98}$$

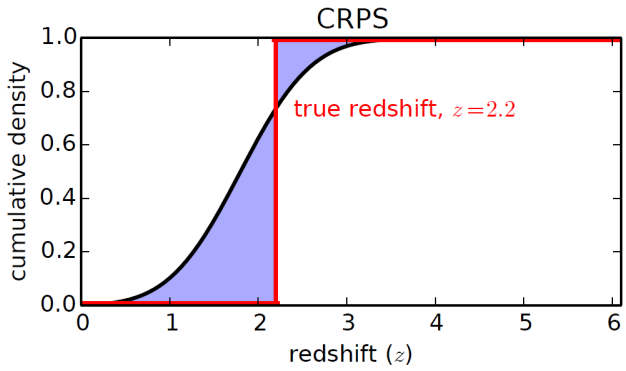

Figure 20: An illustration of the CRPS. assuming the true redshift of an object to be $z_{\rm spec} = 2.2$, integrating the function $C(z) - \mathcal{H}(z - z_{\rm spec})$ would give the purple shaded areas. However since we are integrating the square of this function, we would get an area almost half of this instead. Figure obtained from Polsterer et al. (2016).

where n is the number of bins the histogram is plotted. In this work we bin $Q_{\rm data}$ in number of bins equivalent to half the number of objects in the testing sample. This essentially measures how close the QQ curve is to the diagonal line: the smaller the value, the better the overall PDF shape;

2. *PDF bias* $\rho_{\rm bias}$,

$$\rho_{\rm bias} = \frac{1}{n}\sum(Q_{\rm theory} - Q_{\rm data}), \tag{99}$$

which measures how far the QQ curve goes below the diagonal line, proportional to the area between the curve and the diagonal line. This is a measure of over-prediction of the PDFs, or under-prediction if the values are negative;

3. *PDF wideness* $\rho_{\rm wide}$,

$$\rho_{\rm wide} = \frac{1}{2n}\left[\sum_{Q_{\rm theory}<50\%}(Q_{\rm theory} - Q_{\rm data}) - \sum_{Q_{\rm theory}>50\%}(Q_{\rm theory} - Q_{\rm data})\right], \tag{100}$$

which measures if on average the PDFs are wider (positive) or narrower (negative) than expected. Essentially the QQ curves with 'Z' shapes have positive values, while 'S' shapes have negative values.

Lastly, we also have the *continuous rank probability score* (CRPS), it is a measure of forecast accuracy, telling us how accurate a $p(z)$ predicts the redshift. The CRPS is the squared area between the 'forecast' and the 'observation', whereby in our context, the 'forecast' is the CDF $C(z) = \int_0^z p(z)\,dz$, while the 'observation' is the spec-z, which can be modelled via a Heaviside function $\mathcal{H}(z - z_{\rm spec})$. Mathematically, the CRPS for a single galaxy is

$$CRPS = \int_{-\infty}^{\infty} \left|C(z) - \mathcal{H}(z - z_{\rm spec})\right|^2 dz. \tag{101}$$

To understand this graphically, the difference between $C(z)$ and $\mathcal{H}(z - z_{\rm spec})$ is the purple shaded area in Figure 20, since the integration is the square of the difference of these functions, this results in a smaller area than that since the values are less than 1. The CRPS takes the same unit as redshift, the lower the score the better the overall $p(z)$ captures the spec-z within its curve. It can be used to evaluate the PDFs individually, while for an overall performance of the $p(z)$ we take an average value of the CRPS in the sample, which we denote $\rho_{\rm CRPS}$ in this work.

5.7.3 *Redshift Distributions*

In this work we introduce 2 metrics to quantify photo-z distributions. The first is the *Kolmogorov-Smirnov (KS) statistic* $n_{\rm KS}$, it is the maximum difference between $F_{\rm phot}(z)$ and $F_{\rm spec}(z)$, the CDF of the photometric and spectroscopic redshift distributions $n_{\rm phot}(z)$ and $n_{\rm spec}(z)$ respectively[2],

$$n_{\rm KS} = \max\left(\left|F_{\rm phot}(z) - F_{\rm spec}(z)\right|\right). \tag{102}$$

The KS test is defined with a null hypothesis that the two redshift distributions are drawn from a same distribution, independent of binning. This means that the lower the $n_{\rm KS}$ value, the better the photo-z distribution fits the spec-z distribution. The null hypothesis is rejected at 5% level if

$$n_{\rm KS} > 1.36\sqrt{\frac{2}{N}} \tag{103}$$

where N is the sample size of the testing set. As the KS statistic is distance measurement, we expect it to be sensitive to the maximum differences between distributions. For the case of ANNZ2, BPZ and DELIGHT, $F_{\rm phot}(z)$ is calculated by stacking the PDFs of every single object in the sample, producing a smooth cumulative $n(z)$ to be compared with $F_{\rm spec}(z)$. Since ANNZ does not produce PDFs, the $F_{\rm phot}(z)$ and $F_{\rm spec}(z)$ are simply the cumulative distributions of $z_{\rm phot}$ and $z_{\rm spec}$ respectively.

The final metric is just a simple *root-mean-square difference between the distributions*, denoted as $n_{\rm RMS}$,

$$n_{\rm RMS} = \sqrt{\int \left(n_{\rm spec}(z) - n_{\rm phot}(z)\right)^2 dz}. \tag{104}$$

This quantifies on average how close the photo-z distribution is to the original spec-z distribution.

2 Not to be confused with $C(z)$, the CDF of the $p(z)$ of individual galaxies.

6

PROJECT 1: MORPHOLOGY V.S. PHOTOMETRIC REDSHIFTS

He spreads out the northern skies over empty space; he suspends the earth over nothing.

Job 26:7

6.1 INTRODUCTION

One of the advantages of machine learning photo-z's is that they are capable of incorporating extra inputs on top of the usual broadband magnitudes into training to obtain better results. As we have seen in Section 3.3.6, many attempts have been made to include galaxy morphology as extra inputs to improve photo-z's. Prominent recent analyses in this area include the work by Way (2011), Jones and Singal (2017) and Gomes et al. (2018). However, many of these studies are strongly dependent on the sample the analyses are conducted on. Besides, the magnitude of impact and the performance metrics vary from one another.

Therefore this research is motivated to conduct a thorough study on the impact of galaxy morphology on photo-z's: using the machine learning codes ANNZ and ANNZ2. Galaxy morphology will be included in a large variety of different galaxy samples and surveys, ranging from the bright and low redshift samples up to samples as deep as $i \sim 24.5$ and $z_{\rm spec} \sim 2$. In the process, photometry and morphology from different samples with different qualities will be used, and spectroscopic redshifts from different samples are utilised as well. Most important of all, the photo-z performance of all these samples will be evaluated using the same metrics: namely the root-mean square error ($\sigma_{\rm RMS}$) to evaluate the scatter, the 68th percentile error (σ_{68}) to evaluate the tightness of fit, and the outlier fraction rate ($\eta_{\rm out}$) to evaluate the number of catastrophic outliers in the analyses (definitions described in Section 5.7.1).

This chapter is divided into 4 main sections based on the 4 different types of galaxy samples used:

1. The MGS, a bright SDSS sample reaching $i \sim 17.5$ and $z_{\rm spec} \sim 0.3$ where we study the impact of individual morphological parameters on photo-z's (Section 6.2.1), the effects of morphology on photo-z's separately in elliptical and spiral galaxy samples (Section 6.2.2) and the effect of the spectroscopic or photometric fundamental plane on photo-z's;

2. Colour selected samples (LRG, CMASS and LOWZ) reaching $i \sim 20.5$ and $z_{\rm spec} \sim 0.8$ where a brief analysis on the impact of size on these samples are conducted (Section 6.3);

3. The Stripe-82 Sample, a sample reaching $i \sim 24$ and $z_{\rm spec} \sim 1.5$ where we study the impact of morphology on photo-z's in bright and faint subsamples (Section 6.4.1), samples with a mixture of galaxies and quasars as a result of an imperfect star-galaxy separator (Section 6.4.2) and a sample which includes infrared magnitudes (Section 6.4.3); finally,

4. The CS82 Sample, essentially having the properties as the Stripe-82 Sample, but cross-matched with higher quality galaxy morphology and reweighted in magnitude to the full CS82 Survey (also known as the CS82 Target Sample in this work). In this sample we once again study the impact of individual morphological parameters on photo-z's (Section 6.5.1), and also the impact of morphology on photo-z's when the number of filters are limited (Section 6.5.2) or when the quality of photometry is low (Section 6.5.3).

The results of Section 6.4.2 and Section 6.5 have been published in Soo et al. (2018). All the sources of photometry, spectroscopic redshifts and samples used in this chapter are described in Section 4.1 and Section 4.3.

6.2 THE MAIN GALAXY SAMPLE

6.2.1 *Improvement by Morphological Parameter*

To kick-start our analysis on morphology v.s. photo-z, the first question we would like to tackle, is to know which morphological parameter would improve photo-z's the most. For this analysis we choose to start with the MGS, a relatively bright sample with good photometry to work with. We obtain MGS photometry and spectroscopic redshifts from the SDSS CAS, and with it we also gather a list of available morphological parameters on the `PHOTOOBJ` catalogue: the de Vaucouleurs[1] radius r, Petrosian 50% and 90% light radius $r_{\rm P50}$ and $r_{\rm P90}$, de Vaucouleurs axis-ratio q, Stokes Q and U parameters and the fraction of de Vaucouleurs fit f. As these morphological parameters are observed in all bands, we only chose those observed in the i-band as they have smaller errors than those in other bands.

In addition to these parameters we also derive a few other morphological parameters: the circularised radius $s = r\sqrt{q}$ (also known as 'size' throughout this work); the de Vaucouleurs surface brightness $\mu = i_{\rm deV} + 2.5\log(2\pi q r^2)$, where $i_{\rm deV}$ is the de Vaucouleurs i-band magnitude; the concentration index $C = \frac{r_{\rm P90}}{r_{\rm P50}}$, and the shape probability parameter $p = \frac{\chi^2_{\rm exp}}{\chi^2_{\rm exp}+\chi^2_{\rm deV}}$, derived from the χ^2 values of the model fits. We also cross-matched the MGS with the NYU-VAGC to obtain the Sérsic index for each galaxy. Therefore this gives us 12 morphological parameters to be tested in this work. The reader could refer to Section 2.2.2 for more details on these morphological parameters.

After obtaining all the above-mentioned photometry, morphological parameters, redshifts and removing outliers and bad data, we have a spectroscopic sample of 651 926 objects, where we

1 Throughout the next few sections we have chosen to use the de Vaucouleurs fit radius and axis-ratio as they are the best fit parameters for elliptical galaxies, which are more abundant at low-redshift. However, we later switch to use the exponential radius and axis-ratio for the CS82 Sample in Section 6.5, since it is a much deeper sample. We therefore drop subscripts 'deV' or 'exp' every time we use a radius or axis-ratio for better presentation. Having said that, we note that we have tested radii from both models in all cases and assure the reader that the differences between them are in fact marginal, leaving the conclusion of our analysis the same in all cases.

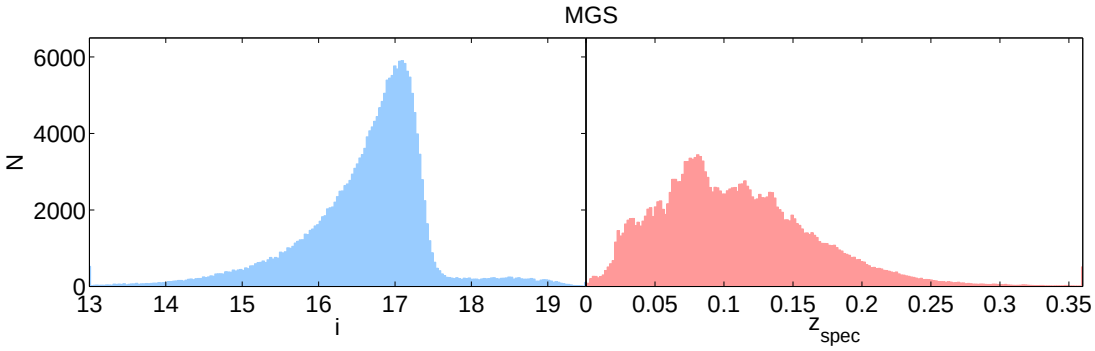

Figure 21: Distributions of the i-band (left) and $z_{\rm spec}$ (right) for the testing set of the MGS.

divide the sample equally into three portions for training, validating and testing. The distribution of the i-band and $z_{\rm spec}$ for the testing set is shown in Figure 21.

Using ANNZ, we first obtain photo-z's for the testing set by training with only *ugriz* photometry, then in subsequent trainings we add morphological parameters to train with *ugriz*, one parameter at a time (the exception being Stokes Q and U parameters, since they describe ellipticity collectively, see Section 2.2.2 for more details). The results of these different trainings are summarised in Table 9. We measure the change in percentage of these metrics with respect to the training without morphology and show them in the table as well.

From the table, we can analyse the results by grouping the morphological parameters into 4 groups: surface brightness (μ), radial measurements (r, s, $r_{\rm P50}$ and $r_{\rm P90}$), ellipticity measurements (q and QU) and shape measurements (C, n, f and p). From the table, we see that μ brings about the most improvement to photo-z's, reaching an improvement in $\sigma_{\rm RMS}$ as high as 3.7%. We expected surface brightness to carry redshift information since theoretically it has a $\log(1+z_{\rm spec})^4$ dependence with redshift. This is followed by radial measurements, for which we also have expected to bring high improvement since it correlates with redshifts through the angular diameter distance. Here we find that the circularised radius actually brings more improvement than the radius itself, this however is mainly due to an artefact in the distribution of the de Vaucouleurs and exponential radii specific to SDSS: spikes at regular radius intervals are seen in the distribution (Figure 22). We have not found any literature addressing this issue, however we suspect it to be a result of the model fitting algorithm.

The initial motivation to use s was to test the difference between the impact of radius with and without circularisation, however due to this artefact in SDSS we are unable to do so, and in subsequent analyses we choose s over r as our primary radial measurement mainly to avoid this artefact. We also see that $r_{\rm P90}$ does not improve as much as $r_{\rm P50}$, however when trained together (second last line of Table 9) provides higher improvement than size itself.

From Table 9 we also notice that shape and ellipticity measurements do not improve photo-z's as much as radial measurements, although q brings improvement comparable to size. This is expected, in fact we could deduce this results based on the correlation between these parameters with redshift. As seen in Figure 23, we see a higher correlation between redshift and surface brightness (top panels) than concentration and axis-ratio. For the case of size, we expect galaxies further from us (up to a certain redshift) to have smaller apparent sizes and vice versa,

Table 9: Relative improvement (Δ%) by morphological parameter in root-mean-square error ($\sigma_{\rm RMS}$), 68th percentile error (σ_{68}) and outlier fraction ($\eta_{\rm out}$) for the MGS, with respect to training with only *ugriz*. The definition of these morphological parameters can be found in Section 2.2.2.

Input vars. (*ugriz*+)	$\sigma_{\rm RMS}$	Δ%	σ_{68}	Δ%	$\eta_{\rm out}(\%)$	Δ%
-	0.0206		0.0173		0.0331	
μ	0.0199	3.7	0.0164	5.2	0.0299	9.7
r	0.0203	1.5	0.0169	2.0	0.0341	−2.8
s	0.0199	3.3	0.0164	5.4	0.0327	1.4
$r_{\rm P50}$	0.0202	2.2	0.0167	3.3	0.0327	1.4
$r_{\rm P90}$	0.0205	0.8	0.0170	1.6	0.0350	−5.6
q	0.0200	2.9	0.0166	3.9	0.0354	−6.9
Q, U	0.0203	1.4	0.0169	2.4	0.0322	2.8
C	0.0202	1.8	0.0168	2.7	0.0336	−1.4
n	0.0205	0.5	0.0171	1.2	0.0359	−8.3
f	0.0204	0.9	0.0171	1.2	0.0285	13.9
p	0.0204	0.9	0.0170	1.7	0.0341	−2.8
r, q	0.0190	8.0	0.0155	10.5	0.0313	5.6
$r_{\rm P50}, r_{\rm P90}$	0.0195	5.4	0.0161	7.1	0.0308	6.9
μ, r, q, n, p	0.0181	12.4	0.0146	15.6	0.0285	13.9

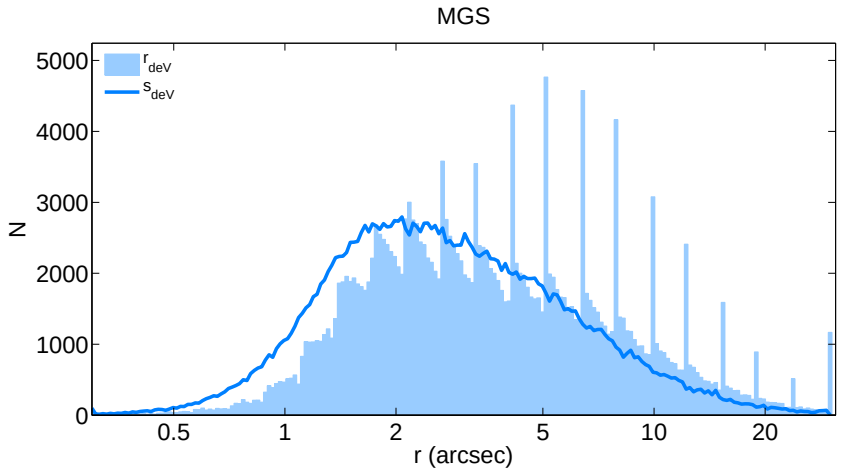

Figure 22: Distributions of the de Vaucouleurs radius $r_{\rm deV}$ (blue histogram) and its circularised radius $s_{\rm deV}$ (blue line) in the MGS. Periodic spikes are seen in the distribution of $r_{\rm deV}$.

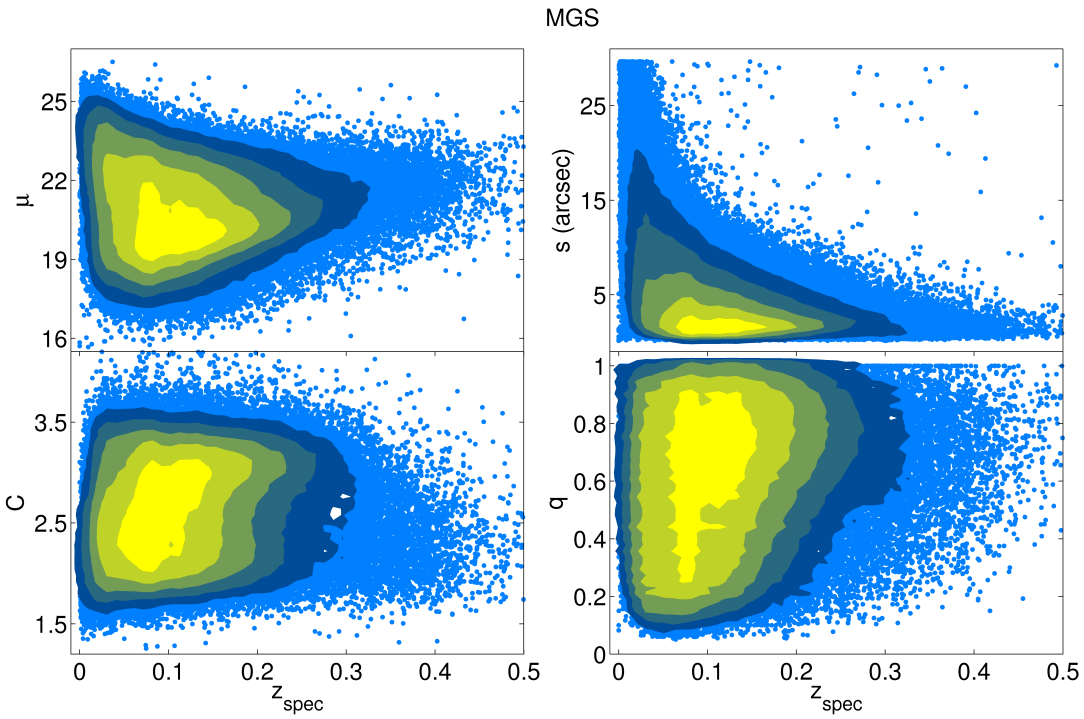

Figure 23: Correlation between surface brightness μ, size s, concentration index C and axis-ratio q with $z_{\rm spec}$ for the MGS.

and the ANN use this information to break degeneracies in colour when training. The training with f may have a high improvement in outlier fraction, however since the outlier fraction for this sample is exceptionally small ($\approx 0.03\%$), $\eta_{\rm out}$ is not a good metric to evaluate the photo-z performance for this particular sample.

We also tried to train multiple morphological parameters with *ugriz*, and from Table 9 we find that training *ugriz* with the ingredients of certain parameters tend to perform better, i.e. training with both r and q compared to s, and training both $r_{\rm P90}$ and $r_{\rm P50}$ compared to C, in both cases we see greater improvement. This means that the improvement s and C individually bring to improve the photo-z does not surpass the improvement brought by the intrinsic correlation the ANN finds with their respective ingredient parameters. Finally we train with *ugriz* and 5 morphological parameters (μ, r, q, n and p) and we find it bringing the most improvement, improving the σ_{68} as much as 15.6% as shown in Table 9, and visualised in Figure 24. The improvement morphology brings to photo-z's is significant for the MGS.

We make a special note here that the choice of using the combination of μ, r, q, n and p were based on the top improving morphological parameters in this sample, and also accounting for their availability in different surveys for easier comparison. We also acknowledge that μ is in fact a derived quantity of $i_{\rm deV}$, r and q, where r and q have technically been used in repetition in this case. We believe that the results could be improved by doing a PCA on all the morphological parameters, but we deem it unnecessary for the purpose of this study as the photo-z results using $\mu rqnp$ is significant and satisfactory. We shall leave the use of PCA in this study as a possibly future work.

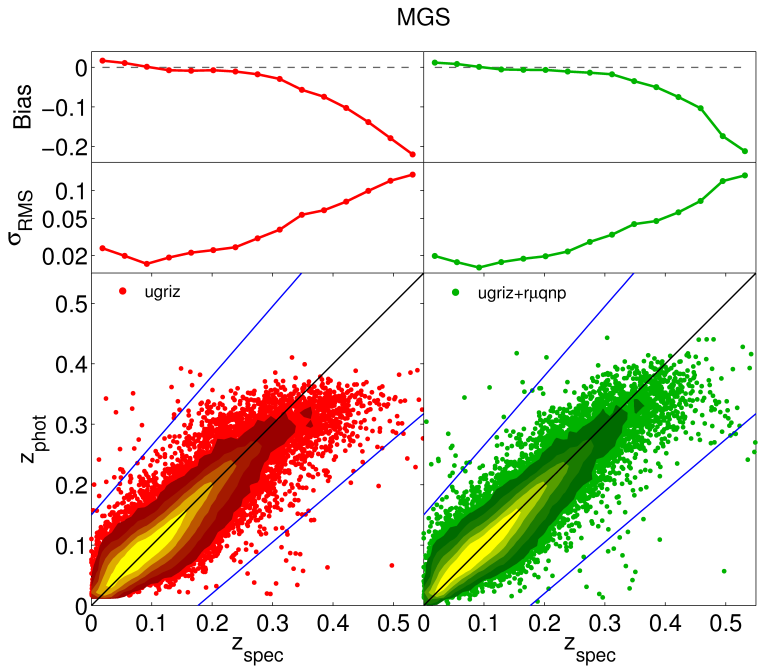

Figure 24: Photo-z v.s. spec-z of the MGS testing set, comparing results between training with *ugriz* (left), and training with *ugriz* and radius r, surface brightness μ, axis-ratio q, Sérsic index n and shape probability p. The coloured contours depict the density of objects, and the blue lines are the limits of the outlier fraction (these lines will be seen in all photo-z v.s. spec-z plots throughout this thesis). The top two panels show the bias and scatter ($\sigma_{\rm RMS}$) as they vary across $z_{\rm spec}$. It is notably visible that the points have moved closer to the diagonal when multiple morphological parameters are included in training, and the bias at the higher redshift end is corrected.

Therefore we conclude that for a bright sample like the MGS, surface brightness and size bring more improvement to photo-z's compared to ellipticity and shape measurements, and the use of multiple morphological parameters brings about even more improvement to photo-z's. However we do not assume that adding more parameters means an improved photo-z, since adding extra irrelevant parameters may contribute to noise on top of an increased dimensionality in parameter space. We take that $\mu rqnp$ are a good combination of morphological parameters to be used collectively to improve photo-z's, and these morphological parameters will be used throughout the next few sections.

6.2.2 *Photometric Redshifts in Elliptical and Spiral Galaxy Samples*

Working on the same sample, we were interested to know if the same results would be seen separately on subsamples of elliptical and spiral galaxies. or more specifically: does galaxy morphology improve photo-z's the same way in elliptical galaxies compared to spiral galaxies? This analysis is motivated by the work of Way (2011) who also studied the effects of galaxy morphology on photo-z's in the context of MGS, however we intend to focus more on the the individual morphological parameters themselves.

The first problem one would encounter however is a reliable way to separate the ellipticals and spirals. The de Vaucouleurs and exponential profile fits have been widely used as proxies to distinguish the galaxy type, e.g. the higher the value of f (`fracDeV`), the more probable a galaxy has a de Vaucouleurs fit and therefore being elliptical, and the lower the value of p which is based on the relative χ^2 value of the de Vaucouleurs fit, the higher the probability the object being an elliptical galaxy. Other morphological parameters are used to distinguish the galaxy type too: when $n \approx 1$ or $C \approx 3.33$, it indicates that the galaxy is probably a disc galaxy, while $n \approx 4$ or $C \approx 2.32$ indicates the galaxy is probably an elliptical galaxy.

In this study however we prefer to approach the galaxy-type separation process independent of the model fit, and this is done with the help of Galaxy Zoo (Banerji et al. 2010), which has been described in Section 4.1.1. In Galaxy Zoo, the galaxy types for galaxies in the MGS are identified by eye: participants are given pictures of galaxies and are asked to indicate if the galaxy is elliptical or if clockwise /anticlockwise spiral features are present. The answers of participants are stored as 'votes': the number of times each galaxy being classified as a certain type by different participants is stored and votes (or probabilities) are output based on the relative number of counts.

Therefore we divide the MGS into elliptical and spiral subsamples as follows. We first cross-match the MGS with the Galaxy Zoo data (the `ZOOVOTES` table in CAS), we then classify an object by its type if it has at least 70% vote of that type, i.e. objects are ellipticals if they have `P_EL>0.7` and combined (clockwise or anticlockwise) spirals if `P_CS>0.7`. Objects with probabilities less than 70%, or with less than 10 single votes are discarded and deemed unreliable.

Here we note that ideally we would have separated the samples as elliptical and disc galaxies (since spirals and lenticulars are subsets of discs), however due to the instructions of Galaxy Zoo not mentioning the word 'disc', most lenticular galaxies have ended up being classified as unknown, and some classified as ellipticals with low vote counts (Lintott et al. 2008). Besides, we expect the number of lenticular galaxies to be outnumbered by spiral galaxies, therefore we

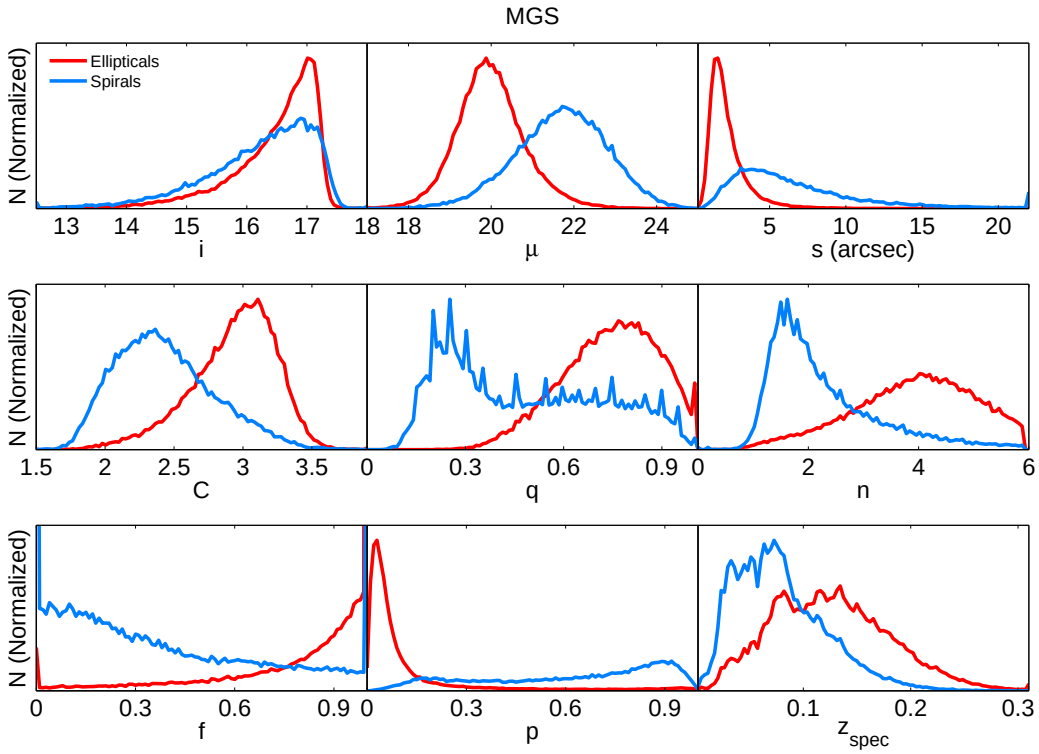

Figure 25: Distributions of the i-band, μ, s, C, q, n, fraction of de Vaucouleurs fit f, p and z_{spec} for the elliptical (red) and spiral (blue) galaxy subsamples of the MGS.

justify that the separation of ellipticals and spirals to be a reasonable proxy for a separation of ellipticals and discs.

After separating the galaxies, we gain a subsample of 197 865 elliptical galaxies and 107 431 spiral galaxies, both divided equally for training, validating and testing. We tabulate the distribution of the i-band, the morphological parameters and redshift for both samples in Figure 25, and we find that although they share the same distribution for i, they both have distinct distributions for every other morphological parameter. We see that the distribution of C, n and p are shown to be true to their expected distribution, but from the distribution of f we see a non-negligible amount of spiral galaxies having a preferred de Vaucouleurs fit.

Using ANNZ, we repeat the exercise of Section 6.2.1 by training and testing both sets with *ugriz* only, then *ugriz* with individual and multiple morphological parameters, we tabulate the results for the elliptical and spiral subsamples in Table 10 and Table 11. Overall we see that when photo-z's are estimated separately for both the elliptical and spiral subsamples, they yield lower σ_{RMS}, σ_{68} and η_{out} when compared to the combined MGS, essentially having almost no outliers. At first glance, this may appear to be a result of sample selection, since by keeping galaxies with high probability measurements may have removed galaxies with lower signal-to-noise. However we have cross-matched galaxies from these two subsamples with galaxies in the testing set of Section 6.2.1, and when the photo-z's of these overlapping galaxies are compared, we find that the photo-z's obtained by training in the elliptical and spiral subsamples perform better than when trained combined: an improvement between 6% to 9% in scatter and σ_{68} is observed, re-

Table 10: Relative improvement by morphological parameter in $\sigma_{\rm RMS}$, σ_{68} and $\eta_{\rm out}$ for the MGS elliptical galaxy subsample with respect to training with only *ugriz*.

Input vars. (*ugriz*+)	$\sigma_{\rm RMS}$	Δ%	σ_{68}	Δ%	$\eta_{\rm out}(\%)$
-	0.0145		0.0119		0.0061
μ	0.0139	4.3	0.0114	4.4	0.0015
r	0.0140	3.4	0.0114	4.4	0.0045
s	0.0140	3.4	0.0114	4.4	0.0030
$r_{\rm P50}$	0.0141	2.9	0.0115	3.3	0.0030
$r_{\rm P90}$	0.0141	2.9	0.0115	3.3	0.0015
q	0.0142	2.2	0.0117	1.5	0.0045
Q, U	0.0142	2.2	0.0117	1.5	0.0045
C	0.0142	2.2	0.0117	1.5	0.0045
n	0.0142	2.2	0.0118	1.1	0.0045
f	0.0142	2.2	0.0118	1.1	0.0045
p	0.0143	1.3	0.0118	1.1	0.0045
r, q	0.0137	5.3	0.0112	5.5	0.0015
$r_{\rm P50}, r_{\rm P90}$	0.0137	5.3	0.0113	5.1	0.0030
μ, r, q, n, p	0.0132	9.0	0.0109	8.7	0.0015

Table 11: Relative improvement by morphological parameter in $\sigma_{\rm RMS}$, σ_{68} and $\eta_{\rm out}$ for the MGS spiral galaxy subsample with respect to training with only *ugriz*.

Input vars. (*ugriz*+)	$\sigma_{\rm RMS}$	Δ%	σ_{68}	Δ%	$\eta_{\rm out}(\%)$
-	0.0191		0.0172		0.0056
μ	0.0185	3.1	0.0165	4.4	0.0056
r	0.0183	4.0	0.0165	4.6	0.0028
s	0.0186	2.8	0.0166	3.6	0.0000
$r_{\rm P50}$	0.0186	2.8	0.0167	3.3	0.0000
$r_{\rm P90}$	0.0185	3.0	0.0167	3.3	0.0056
q	0.0163	14.8	0.0145	16.1	0.0084
Q, U	0.0170	11.1	0.0151	12.2	0.0084
C	0.0184	3.8	0.0165	4.4	0.0056
n	0.0187	2.2	0.0167	3.1	0.0056
f	0.0188	1.7	0.0168	2.8	0.0084
p	0.0185	2.9	0.0166	3.7	0.0084
r, q	0.0156	18.1	0.0139	19.7	0.0028
$r_{\rm P50}, r_{\rm P90}$	0.0177	7.4	0.0159	8.0	0.0000
μ, r, q, n, p	0.0147	22.9	0.0131	24.1	0.0028

gardless of whether trained with morphology. It was also found that the combined training of MGS with 5 morphological parameters (μ, r, q, n, p) performed as good as when trained separately as ellipticals and spirals without morphology. Therefore the separate trainings by galaxy shapes themselves have improved the photo-z's, and the addition of morphology improves them further.

In both subsamples we see a similar trend whereby the photo-z performance is best when more morphological parameters are included in the training. However, the impact of morphology between both samples do differ. Firstly we see that the percentage improvement when morphological parameters are added in training is smaller in the ellipticals when compared to the spirals, agreeing with the results of Way (2011). For the elliptical sample, although we still see more improvement due to surface brightness and radial measurements, the improvement each morphological parameter bring to the photo-z scatter are pretty much almost equal across all parameters, keeping the σ_{RMS} between 0.0140 and 0.0143. This may mainly be due to the fact that with just magnitude information the ellipticals have very good photo-z's to begin with, therefore the changes are just minimal.

On the other hand, we see that the improvement morphology brings to spiral photo-z's are much higher. In Figure 26 we can clearly see the difference in improvement morphology brings towards ellipticals and spirals. In fact, when comparing Figure 26 and Figure 24, the separation between ellipticals and spirals has allowed us to identify that the spiral galaxies are the ones which give rise to the higher scatter when trained combined, and the scatter decreases when morphology is included to correct the photo-z's of these spirals.

What is surprising is that the ellipticity and shape measurements have a higher impact on photo-z's in the spiral sample than the elliptical sample, we further see that axis-ratio and the Stokes parameters bring more improvement than radial measurements. Referring back to Figure 25, we find that most of the spiral galaxies have very high degrees of ellipticity, which may be an indication of the objects inclination. These highly elliptical spiral galaxies might be edge-on spirals which are more affected by dust extinction, which allowed the axis-ratio to play an important role to improve photo-z specifically for such objects.

Therefore we conclude that photo-z's of spiral galaxy samples would benefit most from the inclusion of galaxy morphology in training, especially ellipticity measurements. While the improvement in elliptical galaxies are less, it still would not harm to have multiple morphological parameters included in training as the impact is still a positive one.

6.2.3 *Photometric Redshifts with the Fundamental Plane*

In Section 1.2.9 we have briefly discussed the *fundamental plane*, which is a relationship between three parameters of elliptical galaxies. As we have established in the previous section, Saulder et al. (2013) too utilised the MGS and Galaxy Zoo to construct an elliptical galaxy sample to fit the fundamental plane. We are therefore motivated by their work to see if the empirical relationship of the fundamental plane improves the photo-z's of galaxies in the MGS.

First and foremost, we shall define the spectroscopic fundamental plane,

$$\log(R_0) = \alpha \log(\sigma_0) + \beta(\mu_0) + \gamma, \tag{105}$$

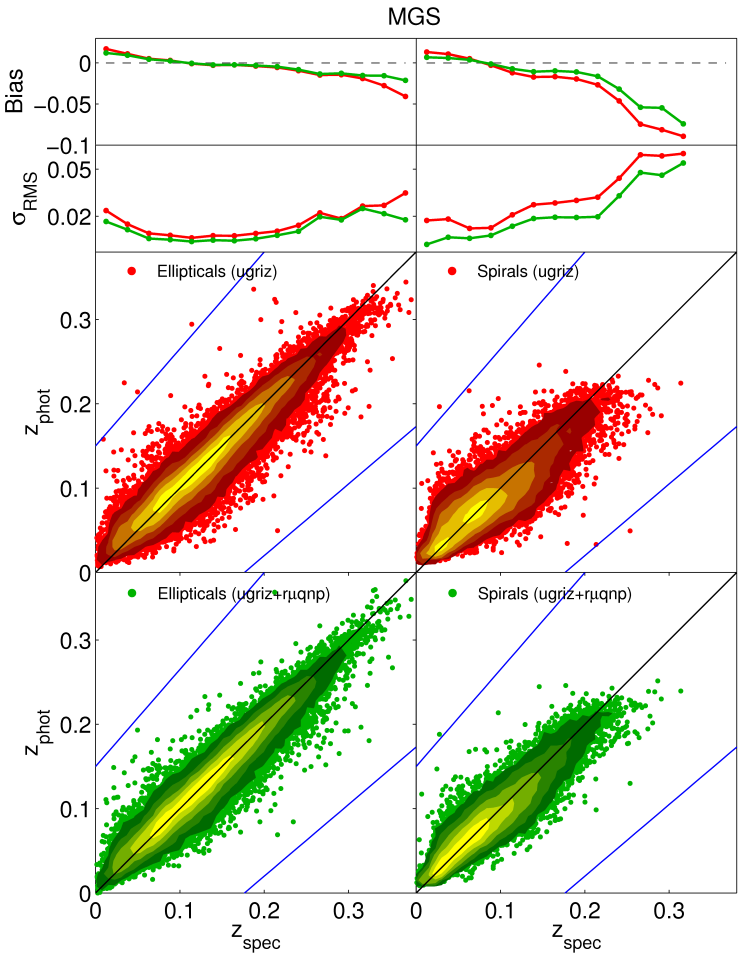

Figure 26: Photo-z v.s. spec-z of the MGS elliptical (left) and spiral (right) galaxy subsamples, comparing results between training with *ugriz* (top) and *ugriz* with morphology (bottom).

where R_0 is the physical effective radius, the σ_0 the velocity dispersion, μ_0 the mean surface brightness and $\{\alpha, \beta, \gamma\}$ are parameters of the fit. Breaking down each of the parameters, the first term R_0 can be written fully as

$$\begin{aligned} R_0 &= d_\mathrm{A}(z_\mathrm{c}) \tan(s) \\ &= \frac{c}{H_0} \frac{z_\mathrm{c}}{(1+z_\mathrm{c})^2} \left[1 + \frac{z_\mathrm{c}(1-q_0))}{\sqrt{1+2q_0 z_\mathrm{c} + 1 + q_0 z_\mathrm{c}}} \right] \tan(s), \end{aligned} \tag{106}$$

where d_A is the angular diameter distance (see Section 1.2.8) with the assumptions of ΛCDM cosmology of a flat and accelerating universe, c the speed of light, $H_0 = 70$ km s^{-1} Mpc^{-1} the Hubble constant, $q_0 = -0.55$ the deceleration parameter and $s = r\sqrt{q}$ the size of galaxies as we have defined previously. z_c is the redshift of the galaxy corrected for the motion of our solar system relative to the CMB, which can be written as

$$\begin{aligned} z_\mathrm{c} &= (\vec{z} + \vec{z}_\mathrm{CMB}) \cdot (\text{line of } \vec{\text{sight}}) \\ &= \left[z \begin{pmatrix} \cos b \cos l \\ \cos b \sin l \\ \sin b \end{pmatrix} + \frac{v_\mathrm{CMB}}{c} \begin{pmatrix} \cos b_\mathrm{CMB} \cos l_\mathrm{CMB} \\ \cos b_\mathrm{CMB} \sin l_\mathrm{CMB} \\ \sin b_\mathrm{CMB} \end{pmatrix} \right] \cdot \begin{pmatrix} \cos b \cos l \\ \cos b \sin l \\ \sin b \end{pmatrix}, \end{aligned} \tag{107}$$

where b and l are the galactic coordinates of the galaxy, with the solar system moving in the direction of $b_\mathrm{CMB} = 48.26° \pm 0.03°$ and $l_\mathrm{CMB} = 264.0° \pm 0.1°$ at velocity $v_\mathrm{CMB} = 369.0 \pm 0.9$ km s^{-1}.

The second term σ_0, the velocity dispersion is corrected for the fixed fibre size in SDSS,

$$\sigma_0 = \sigma \left(\frac{8 r_\mathrm{fib}}{s} \right)^{0.04}, \tag{108}$$

where σ is the velocity dispersion value as stored in SDSS (in km s^{-1}), $r_\mathrm{fib} = 1.5''$ is the SDSS fibre radius. σ_0 is on average 10% larger than σ (Saulder et al. 2013).

The final term μ_0 is the mean surface brightness, written as

$$\mu_0 = m - A - K + E z_\mathrm{c} + 2.5 \log(2\pi s^2) - 10 \log(1 + z_\mathrm{c}), \tag{109}$$

where m is the magnitude band, A the extinction correction, K the K-correction, $E = 1.07$ mag per redshift to correct for evolution, and the final term to correct for surface brightness dimming (see Equation 32 for full details).

With the fundamental plane defined, we now divide each of the terms R_0, σ_0, and μ_0 into spectroscopic and photometric parts:

$$\begin{aligned} \log(R_\mathrm{s}) &= \log(D_\mathrm{A}(z_\mathrm{c})) & \log(R_\mathrm{p}) &= \log(\tan(s)) \\ \log(\sigma_\mathrm{s}) &= \log(\sigma_0) & \log(\sigma_\mathrm{p}) &= 0 \\ \mu_\mathrm{s} &= E z_\mathrm{c} - 10 \log(1 + z_\mathrm{c}) & \mu_\mathrm{p} &= m - A - K + 2.5 \log(2\pi s^2) \end{aligned} \tag{110}$$

with the subscripts s and p referring to spectroscopic and photometric, respectively. We have placed K in the photometric part as we utilise the K-correction values from the SDSS `PHOTOZ`

table, which were obtained photometrically with the help of SDSS photo-z's. Next we rearrange the fundamental plane as follows:

$$\begin{aligned} \log(R_0) &= \alpha \log(\sigma_0) + \beta\mu_0 + \gamma \\ \left[\log(R_\mathrm{s}) + \log(R_\mathrm{p})\right] &= \alpha\left[\log(\sigma_\mathrm{s}) + \log(\sigma_\mathrm{p})\right] + \beta(\mu_\mathrm{s} + \mu_\mathrm{p}) + \gamma \\ -\left[\alpha\log(\sigma_\mathrm{s}) + \beta\mu_\mathrm{s} - \log(R_\mathrm{s})\right] &= \left[\alpha\log(\sigma_\mathrm{p}) + \beta\mu_\mathrm{p} - \log(R_\mathrm{p})\right] + \gamma \\ P_\mathrm{s} &= P_\mathrm{p} \end{aligned} \tag{111}$$

where we define P_s and P_p as the spectroscopic and photometric *plane parameters*.

Therefore in order to incorporate the fundamental plane into our photo-z estimation, we add P_p as an extra parameter in our training, we do not add P_s since we assume that we do not have any spectroscopic information of these galaxies in the testing set. The fit parameters $\{\alpha, \beta, \gamma\}$ in P_p are obtained by fitting the training set with the fundamental plane, where we have spectroscopic information. In other words, we are trying to see if the empirical fit of the fundamental plane provide extra information to improve photo-z's, and this extra information is embedded in the parameters $\{\alpha, \beta, \gamma\}$.

However we do not merely stop at the fundamental plane, we also tried several photometric variations of the fundamental plane, which are known as *photometric planes* (short for photometric fundamental planes). These photometric planes are constructed by replacing σ_0 with different parameters. In this work we tried the following four forms of photometric planes:

$$\log(R_0) = \alpha_1 \log(C) + \beta_1\mu_0 + \gamma_1 \tag{112}$$

$$\log(R_0) = \alpha_2 \log(n) + \beta_2\mu_0 + \gamma_2 \tag{113}$$

$$\log(R_0) = \alpha_3 \log(g\text{-}r) + \beta_3\mu_0 + \gamma_3 \tag{114}$$

$$\log(R_0) = \alpha_4 \log(v_\mathrm{phot}) + \beta_4\mu_0 + \gamma_4. \tag{115}$$

The first one replaces σ_0 with the concentration index, this form is motivated by the work of Huff and Graves (2014) who used this to measure magnification of galaxies due to weak lensing. The second uses the Sérsic index, motivated by the work of Graham (2002) who used this equation to fit elliptical galaxies from Virgo and Fornax. The third equation uses the colour g-r, motivated by the work of Wake et al. (2012) who showed that velocity dispersion best correlate with colour. Finally, we call v_phot the 'photometric velocity dispersion' or photo-v, where we photometrically estimate the velocity dispersion using ANNZ, by training *ugriz*, axis-ratio and radius. In these cases we now have $\log(\sigma_\mathrm{s}) = 0$ while $\log(\sigma_\mathrm{p})$ equals to whichever replacement parameter used, since these are all photometric parameters. Therefore we denote $P_C, P_n, P_{g\text{-}r}, P_v$ to represent the photometric counterparts of these photometric planes coefficients and are included in training.

To test this, we explicitly use the sample selection as Saulder et al. (2013), which we call the *Saulder's Sample*. The selection cuts are as follows: redshift range $0 < z_\mathrm{spec} < 0.5$; `ZWARNING=0`; velocity dispersion 100 km s^{-1} $< \sigma <$ 120 km s^{-1}; the spectra's $^S/_N$ having `SNMEDIAN>10`; `CLASS=GALAXY`; Galaxy Zoo elliptical galaxy votes `P_EL>0.8`; number of total votes `NVOTE_TOT>10`; de Vaucouleurs axis-ratio $q > 0.3$ for every band (to exclude very elongated elliptical and lentic-

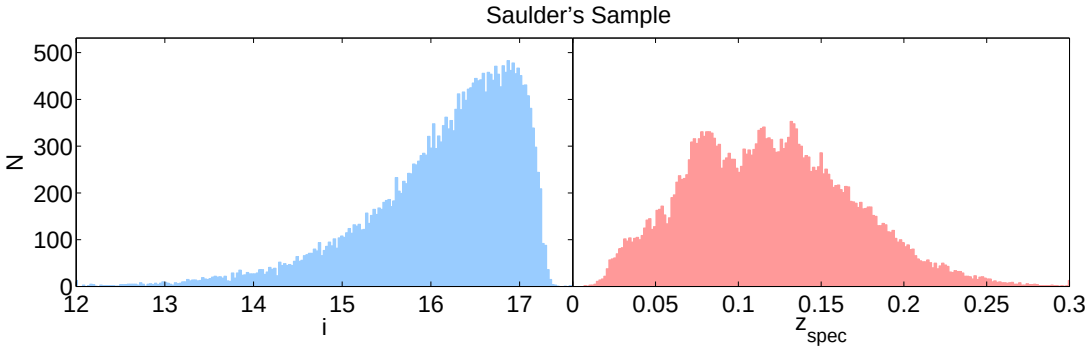

Figure 27: Distributions of the i-band magnitude (left) and $z_{\rm spec}$ (right) for Saulder's Sample.

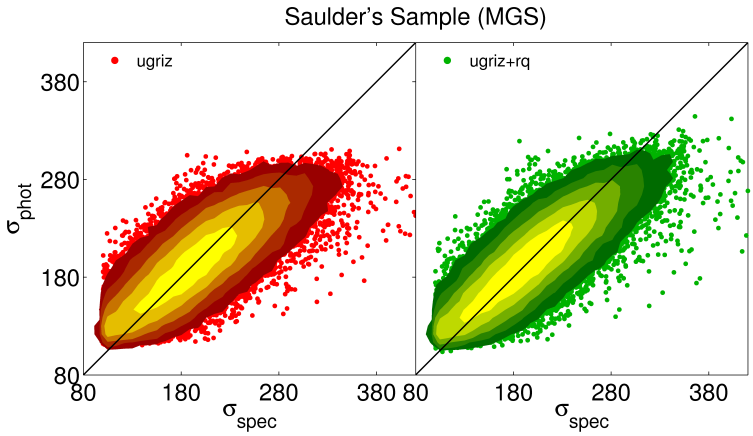

Figure 28: The estimation of 'photometric velocity dispersion' using ANNz, comparing the results trained with $ugriz$ and $ugriz + rq$ (radius and axis-ratio).

ular galaxies); and the log-likelihoods $\log(L_{\rm deV}) > \log(L_{\rm exp})$ for all bands. These cuts yield a sample of 83 104 objects equally divided into three portions for training, validating and testing.

The distributions of i and $z_{\rm spec}$ of the Saulder's Sample are visualised in Figure 27, and it is worth mentioning that this sample is less than half the size of the elliptical sample used previously due to the more stringent cuts. We also show in Figure 28 the photo-v we obtain, utilising the same training and testing set, comparing the training using $ugriz$ and $ugriz$ with axis-ratio and radius, with the latter showing a better fit. The photo-v obtained yield a $\sigma_{\rm RMS}$ of 23.91 km s^{-1}, and this variable is calculated for and used in both the training and testing set to fit the photometric plane P_v.

Using the training set, we fit the coefficients $\{\alpha, \beta, \gamma\}$ of the 5 planes using a least-squares fit, and we summarise the fitted parameters in Table 12. The values of $\{\alpha, \beta, \gamma\}$ are then applied to the respective planes, and we plot $\log(R_0)$ against $\alpha \log(X) + \beta\mu_0 + \gamma$ with X being σ_0 for the fundamental plane or $C, n, g\text{-}r, v_{\rm phot}$ for each photometric plane variation, and this is shown in Figure 29.

Just the fitting of these planes yields some very interesting results. First of all, the two most widely used photometric planes (P_C and P_n) actually yield a lower correlation coefficient than

Table 12: Coefficients $\{\alpha, \beta, \gamma\}$, reduced χ^2 value of the plane fits, and correlation coefficient $r_{\rm corr}$ for the fundamental plane and each of the photometric planes constructed. All errors of the individual parameters are linearly propagated into the fit.

Plane	α	β	γ	χ^2/dof	$r_{\rm corr}$
$P_{\rm p}$	1.01 ± 0.04	0.298 ± 0.007	-4.5 ± 0.2	0.00692	0.9255
P_C	1.2 ± 0.1	0.299 ± 0.007	-2.8 ± 0.2	0.01703	0.8043
P_n	0.235 ± 0.006	0.301 ± 0.001	-2.34 ± 0.02	0.01808	0.7910
$P_{g\text{-}r}$	1.41 ± 0.07	0.257 ± 0.008	-1.3 ± 0.1	0.01080	0.8815
P_v	1.25 ± 0.05	0.295 ± 0.007	-5.0 ± 0.2	0.00570	0.9351

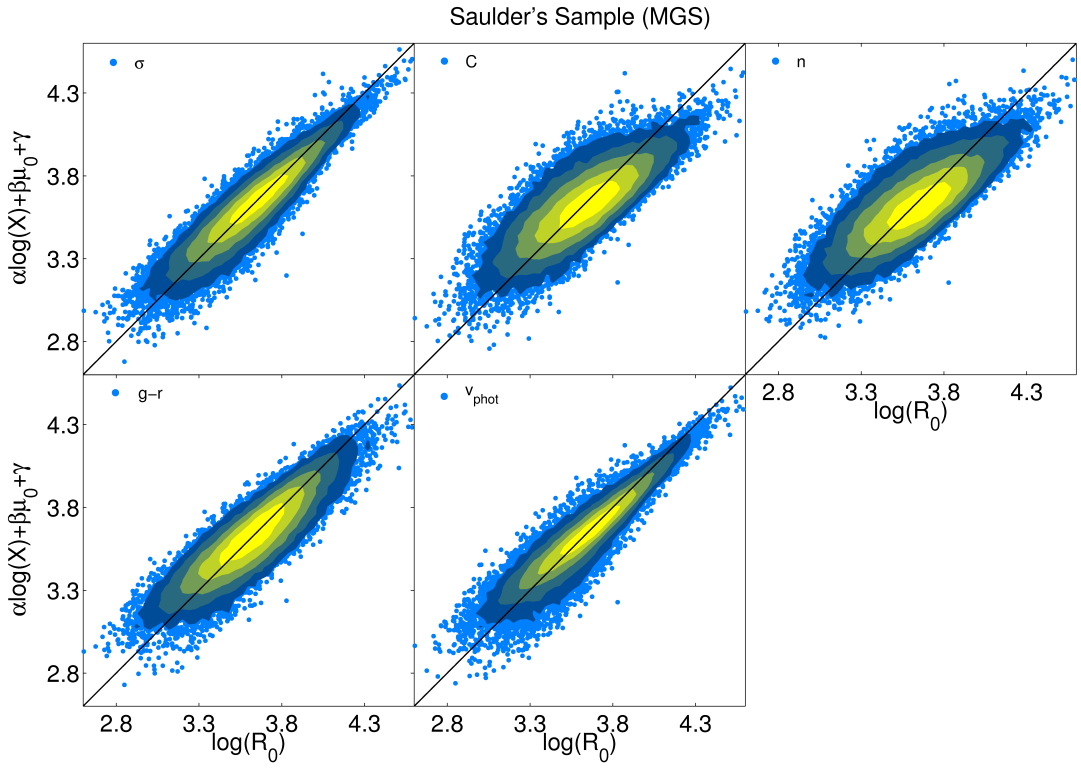

Figure 29: Straight line fit of the spectroscopic fundamental plane (top left), compared to the 4 photometric planes shown in this work, when σ_0 is replaced by C, n, g-r or $v_{\rm phot}$.

Table 13: $\sigma_{\rm RMS}$, σ_{68} and $\eta_{\rm out}$ for galaxies in Saulder's Sample, and their respective improvements when the plane or morphological parameters are added in training when compared to training with only *ugriz*.

Input vars. (*ugriz*+)	$\sigma_{\rm RMS}$	Δ%	σ_{68}	Δ%	$\eta_{\rm out}$(%)
-	0.0110		0.0096		0.0
$P_{\rm p}$	0.0106	3.8	0.0091	4.7	0.0
$P_{\rm C}$	0.0107	2.5	0.0092	4.4	0.0
P_n	0.0107	2.5	0.0093	3.6	0.0
$P_{g\text{-}r}$	0.0106	3.8	0.0092	4.4	0.0
P_v	0.0107	2.5	0.0094	3.0	0.0
$i_{\rm deV}, r, q$	0.0103	6.4	0.0090	6.8	0.0
$i_{\rm deV}, r, q, C$	0.0103	6.4	0.0089	7.5	0.0
$i_{\rm deV}, r, q, n$	0.0103	6.4	0.0089	7.5	0.0
$i_{\rm deV}, r, q, g\text{-}r$	0.0103	6.4	0.0089	7.5	0.0
$i_{\rm deV}, r, q, v_{\rm phot}$	0.0103	6.4	0.0089	7.5	0.0

$P_{g\text{-}r}$. What is more surprising is that P_v actually fits better than the spectroscopic fundamental plane. At this point we are interested to see if the higher correlation coefficient dictates a higher improvement to photo-z, and also how the results would differ if the individual ingredients of the planes were trained instead. We use ANNZ as usual and train these plane parameters and the ingredient morphological parameters with *ugriz*, and the results are summarised in Table 13.

The results shown however are not as exciting as expected: although we see $P_{\rm p}$ and $P_{g\text{-}r}$ performing better than the rest, the low $\sigma_{\rm RMS}$ of 0.01 and zero outliers makes it difficult to judge if this is actual improvement or just fluctuation of noise. But what is definitely clear is that training the individual parameters yield better results than training the plane parameters alone, though it is still possible to distinguish from each plane as to which plane brings about the most improvement. Figure 30 illustrates this, showing that the improvement is practically not visible when $P_{\rm p}$ is added to the training.

Therefore we conclude that although the fundamental and photometric planes are important and useful relations in many situations, they are not helpful for photo-z estimation in the context of MGS, as the photo-z's for a carefully selected sample such as Saulder's Sample has very good photo-z's to begin with. However we believe that more improvement is expected at higher redshift samples, though these samples would face different challenges such as a more reliable galaxy-type separation and also accurate velocity dispersion estimates. We shall leave this to be extended for possible future work.

6.3 COLOUR SELECTED SAMPLES

We now move away from the MGS to see the impact of morphology on deeper and higher redshift samples. We selected a few spectroscopic galaxy samples in the SDSS, namely the LRG, LOWZ and CMASS samples. These samples are colour selected samples, most of them are generally large and red galaxies with the exception of CMASS which extends a little bluer than the

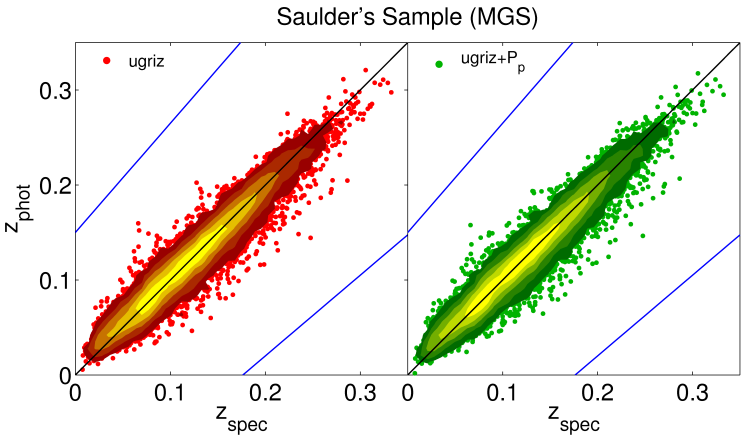

Figure 30: Photo-z v.s. spec-z of Saulder's Sample, comparing results between training of *ugriz* (left) and *ugriz* with the fundamental plane parameter.

Table 14: Improvement in $\sigma_{\rm RMS}$, σ_{68} and $\eta_{\rm out}$ when size is added to training for the 2SLAQ, LRG, LOWZ and CMASS samples, respectively.

Sample	**Input**	$\sigma_{\rm RMS}$	$\Delta\%$	σ_{68}	$\Delta\%$	$\eta_{\rm out}(\%)$	$\Delta\%$
2SLAQ	*ugriz*	0.0284		0.0228		0.12	
	$ugriz+s$	0.0279	1.7	0.0222	2.6	0.06	50.0
LRG	*ugriz*	0.0304		0.0237		0.31	
	$ugriz+s$	0.0301	0.9	0.0234	1.1	0.30	1.8
LOWZ	*ugriz*	0.0233		0.0185		0.10	
	$ugriz+s$	0.0228	2.4	0.0177	4.2	0.10	0.0
CMASS	*ugriz*	0.0408		0.0274		1.17	
	$ugriz+s$	0.0400	2.1	0.0268	1.9	1.06	9.2

LRG to form a stellar-mass limited sample with all intrinsic colours. We also use the spectroscopic redshifts from the 2SLAQ LRG Sample (Cannon et al. 2006) just for comparison with the SDSS LRG Sample. The details of these samples have been discussed in Section 4.3.

In this section we attempt to see the impact of galaxy size on the photo-z's of LRG-like samples. We obtain all of the photometry and spectroscopy from SDSS CAS and 2SLAQ, and after removing outliers we have 9718 objects for the 2SLAQ Sample, 882 577 for the CMASS Sample, 327 555 for the LOWZ Sample and 708 013 for the LRG Sample. The distributions of i and $z_{\rm spec}$ for these samples are visualised in Figure 31, where we notice that with the exception of the LOWZ Sample, the other 3 samples have almost identical distributions of i and $z_{\rm spec}$.

Similar to the previous sections, we divide these samples equally into 3 portions for training, validating and testing. The training of *ugriz* is compared to $ugriz+s$ where s is the size of the galaxy as defined before. Using ANNz, the results of these 4 samples are summarised in Table 14 and visualised in Figure 32.

From the table we see that overall size brings the most improvement to photo-z's of the LOWZ Sample, yielding a change in σ_{68} of 4.2%. We expect that the improvement of size to the LRG and CMASS sample photo-z's to be small since these samples have very stringent cuts and show

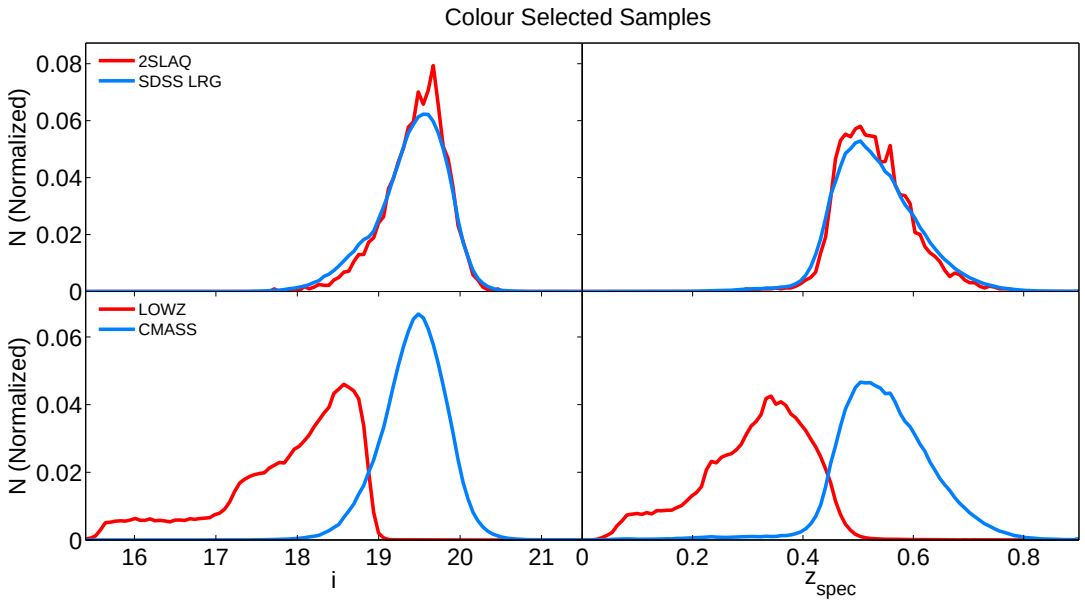

Figure 31: Distributions of the *i*-band (left) and z_{spec} (right) for the 2SLAQ (top red), SDSS LRG (top blue), LOWZ (bottom red) and CMASS (bottom blue) samples, respectively.

small variability in size. From Figure 32 we get to see the signature 'tail' of outliers for both the LRG and CMASS samples (also seen in Abdalla et al. 2011), a flattened upper limit at high redshift can also be seen.

We would like to point out that the fraction of these outliers are in fact very small, as seen from Table 14 (less than 2% of the objects). The large size of the sample and the finite size of the dots give the illusion of a larger scatter. Despite that, it is in fact quite interesting to note that size reduces some of these outliers: from the LRG Sample we see objects of low-redshift getting closer to the diagonal line when trained with size, and the tail of the CMASS Sample moving a little closer too.

The existence of these tails means that there are degeneracies in colour which training with merely *ugriz* was not able to solve. Here, although we have seen that size does improve the photo-*z* for these objects a little, it is evident that additional missing information would be required to further improve this. We have in fact attempted to add concentration and axis-ratio in the training, but the change is even smaller than the use of size. Therefore we conclude that although size does improve the photo-*z*'s of LRG-like samples, the improvement is only significant for the outliers and marginal overall.

6.4 THE STRIPE-82 SAMPLE

6.4.1 *Photometric Redshifts in a Bright and Faint Sample*

Having tested the impact of morphology on photo-*z*'s for various SDSS-based samples, we now move to an even fainter and higher redshift sample, where we have reached the upper limits of SDSS in both photometry and spectroscopy. To test the impact of morphology on samples

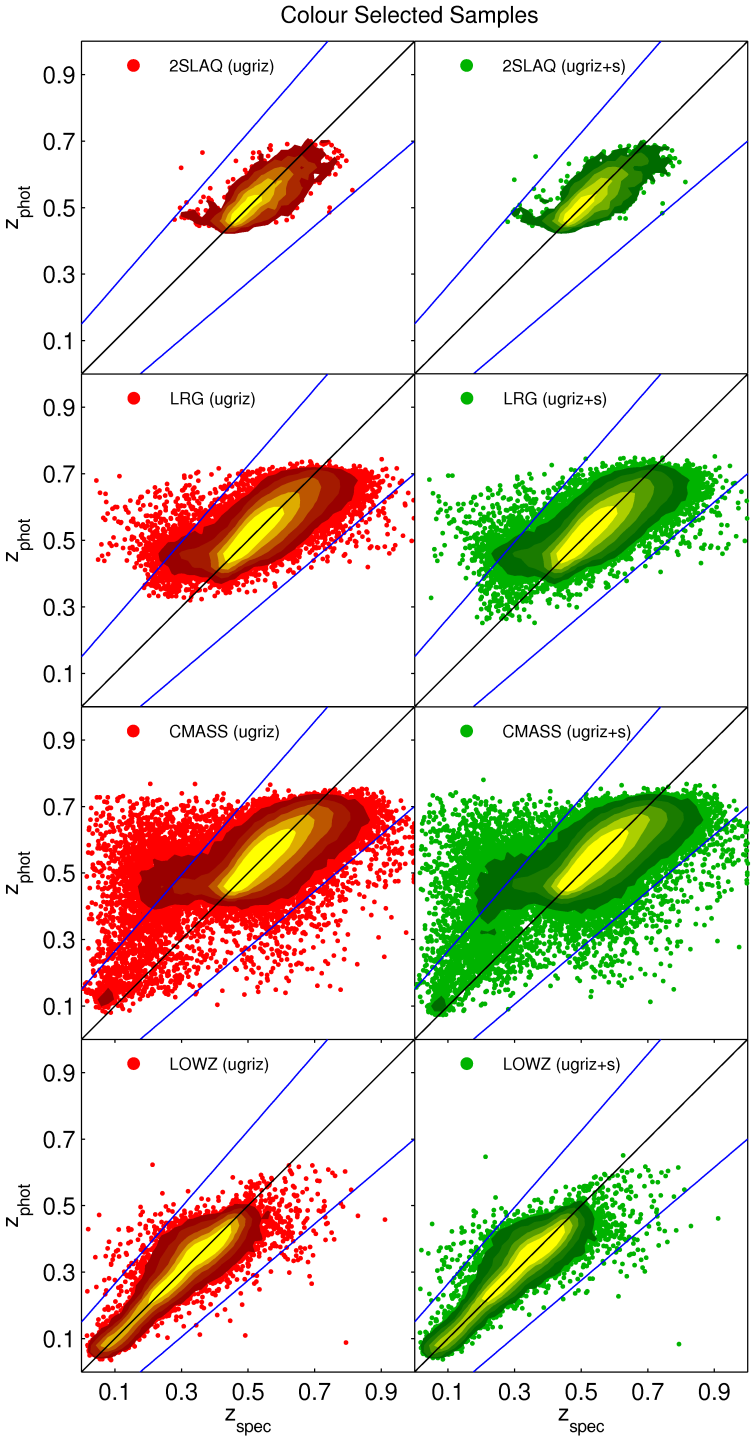

Figure 32: Photo-z v.s. spec-z of the 2SLAQ (first row), LRG (second row), CMASS (third row) and LOWZ (bottom row) samples, comparing the results of *ugriz* (left) and *ugriz* with size (right).

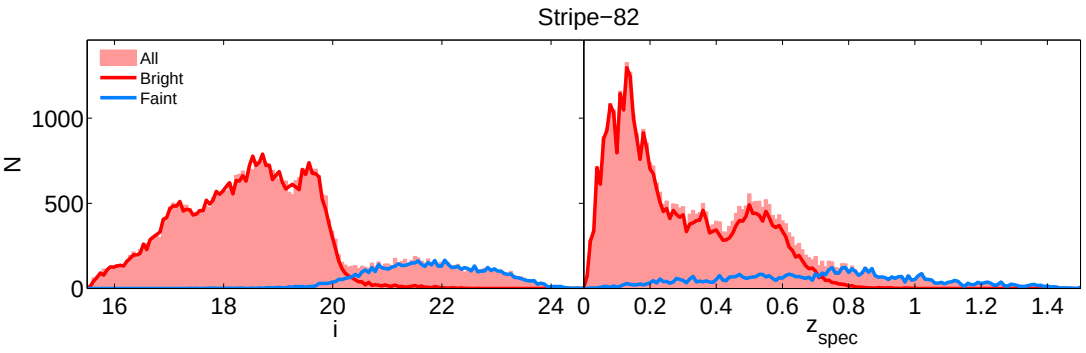

Figure 33: Distributions of the i-band (left) and z_{spec} (right) for the Stripe-82 Sample. The red and blue lines show the distribution of the bright and faint subsamples, respectively.

reaching a magnitude of $i \sim 24$ and $z_{\text{spec}} > 1$, we find the SDSS Stripe-82 Sample suitable for our case. As described in Section 4.1.1, the SDSS Stripe-82 region was observed repeatedly to allow the creation of coadded photometry, reaching at least 2 magnitudes fainter than the standard SDSS observations.

The Stripe-82 region also benefits from the overlap of many photometric and spectroscopic surveys like UKIDSS, DEEP2, WiggleZ and VVDS. To construct a spectroscopic sample, we cross-match the SDSS Stripe-82 coadded photometry with spectroscopic redshifts from SDSS, BOSS, WiggleZ, DEEP2 and VVDS to obtain a sample of 126 154 objects, divided equally for training, validating and testing. The details of these redshift surveys can be found in Section 4.3.

In this sample we only select 3 morphological parameters to work with, namely the surface brightness μ, axis-ratio q and angular radius r. Similar to previous analyses, we use ANNZ to train and compare the photo-z's of the Stripe-82 Sample, with and without several combinations of morphological parameters. On top of that, we also test these results in two subsamples: we select a *bright sample* by using objects which have SDSS or BOSS redshifts, and a *faint sample* by using objects which have DEEP2, WiggleZ and VVDS redshifts. With these cuts we now have a bright and faint subsample with 104 970 and 20 934 objects each. The distributions of i-band magnitude and z_{spec} for the Stripe-82 Sample and its two subsamples are visualised in Figure 33, where we see that the bright and faint samples can be roughly be categorised as objects with $i < 20$ and objects with $i > 20$.

Using ANNZ, Table 15 shows the results for the Stripe-82 Sample, the bright and faint subsamples, respectively. For the entire sample, the familiar trend still applies where surface brightness and size improve photo-z's more than axis-ratio, and multiple morphological parameters improve photo-z's with a higher percentage across all samples. However, when analysing the impact of each individual morphological parameter on the bright sample, the percentage difference for σ_{RMS} and σ_{68} are arguably insignificant, most of the time less than 2% improvement, or experienced degradation. Looking at the faint sample on the other hand, radius seemed to be the driving parameter to improve photo-z's, however the percentage improvement remain low when used alone without other morphological parameters. This suggests that for a general sample of mixed-type galaxies, the improvement each individual morphological parameter bring to the

Table 15: Relative improvement by morphological parameter in $\sigma_{\rm RMS}$, σ_{68} and $\eta_{\rm out}$ for the Stripe-82 Sample, with respect to training with only *ugriz*. Also shown are the results when trained and tested separately on the bright ($i < 20$) and faint ($i > 20$) subsamples.

Sample	**Input vars. (*ugriz*+)**	$\sigma_{\rm RMS}$	$\Delta\%$	σ_{68}	$\Delta\%$	$\eta_{\rm out}(\%)$	$\Delta\%$
All	-	0.0531		0.0286		1.95	
	μ	0.0515	3.1	0.0274	4.4	1.76	9.5
	q	0.0521	2.0	0.0282	1.5	1.80	7.9
	r	0.0520	2.1	0.0270	5.8	1.68	13.7
	s	0.0514	3.2	0.0265	7.6	1.65	15.4
	r, q	0.0509	4.2	0.0250	12.6	1.63	16.3
	μ, r, q	0.0500	5.9	0.0240	16.2	1.58	18.9
Bright	-	0.0298		0.0192		0.51	
	μ	0.0294	1.4	0.0187	2.6	0.48	5.1
	q	0.0298	0.1	0.0190	1.0	0.48	5.6
	r	0.0296	0.7	0.0189	1.4	0.50	1.1
	s	0.0300	-0.6	0.0188	1.9	0.52	-2.8
	r, q	0.0289	2.9	0.0183	4.6	0.47	8.4
	μ, r, q	0.0278	6.6	0.0175	8.7	0.44	14.0
Faint	-	0.0927		0.0639		7.08	
	μ	0.0921	0.6	0.0643	-0.7	6.32	10.7
	q	0.0929	-0.3	0.0642	-0.5	7.14	-0.8
	r	0.0912	1.6	0.0621	2.8	6.56	7.3
	s	0.0935	-0.9	0.0634	0.7	6.39	9.7
	r, q	0.0919	0.8	0.0596	6.7	6.15	13.2
	μ, r, q	0.0905	2.3	0.0591	7.5	6.00	15.2

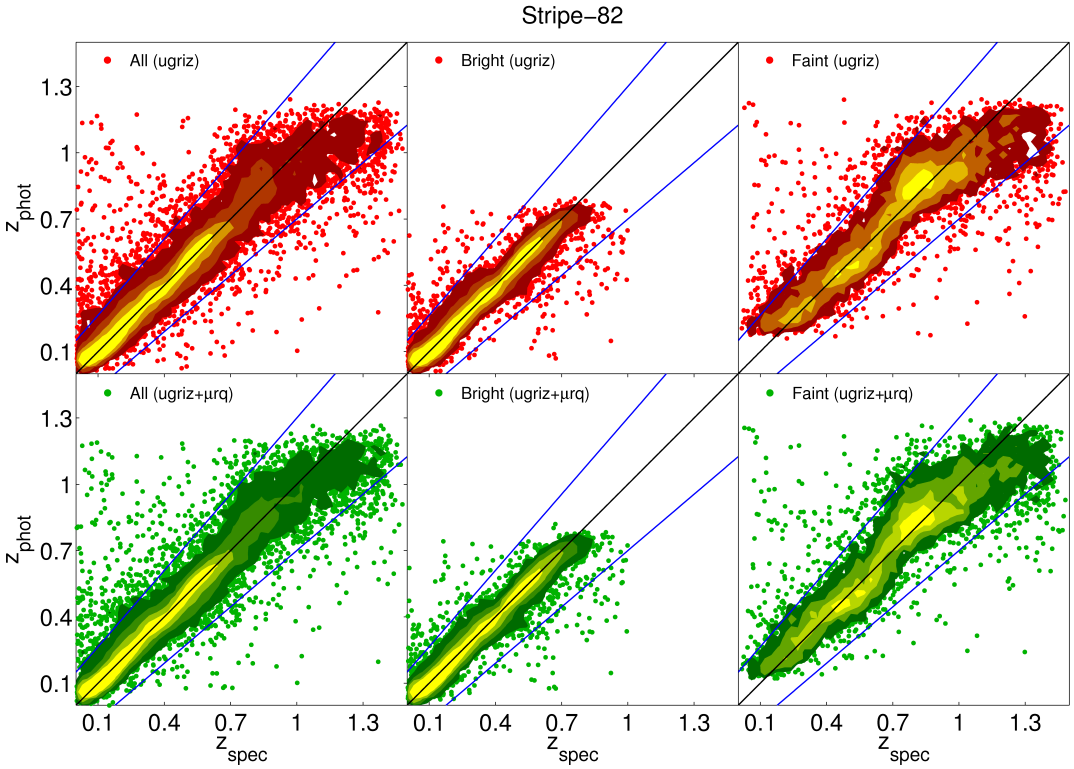

Figure 34: Photo-z v.s. spec-z of the Stripe-82 Sample (left), its bright subsample (middle) and faint subsample (right), comparing the results training with and without morphological parameters μ, r and q (top and bottom).

photo-z is insignificant and comparable to training noise, and visible impact can only be seen if multiple morphological parameters are used.

Figure 34 compares the photo-z trained with only *ugriz* and *ugriz* with multiple morphological parameters for the Stripe-82 Sample, its bright and faint subsets. For the full sample and the faint subsample we see a slight improvement in the high-redshift region, while the improvement for the bright sample is barely visible. We deduce that the improvement or degradation morphology brings to photo-z on a sample reaching $i \sim 24$ are marginal if quantified using global metrics, since the uncertainty of the photo-z's produced are of the same order of magnitude as the impact of morphology.

We attempt to move beyond global metrics of the samples to see if there are any trends of improvement when the metrics σ_{RMS} and σ_{68} are plotted against magnitude, size and redshift. Figure 35 illustrates this for all three samples, and from the left column, we see that galaxy morphology generally improves the σ_{RMS} and σ_{68} more at lower magnitude (the brighter end), we believe this is so since brighter galaxies have larger sizes and lower redshifts, thus the ANN corrects the photo-z's of these objects in the presence of the fainter ones. We do not see any trends as the metric spans across radius when trained for the entire sample, however we see more improvement for larger objects for the bright sample and smaller objects for the faint sample, presumably those are where the bulk of the objects are in those samples. As for z_{spec}, we see more improvement in σ_{68} for the entire sample and bright sample at low redshift.

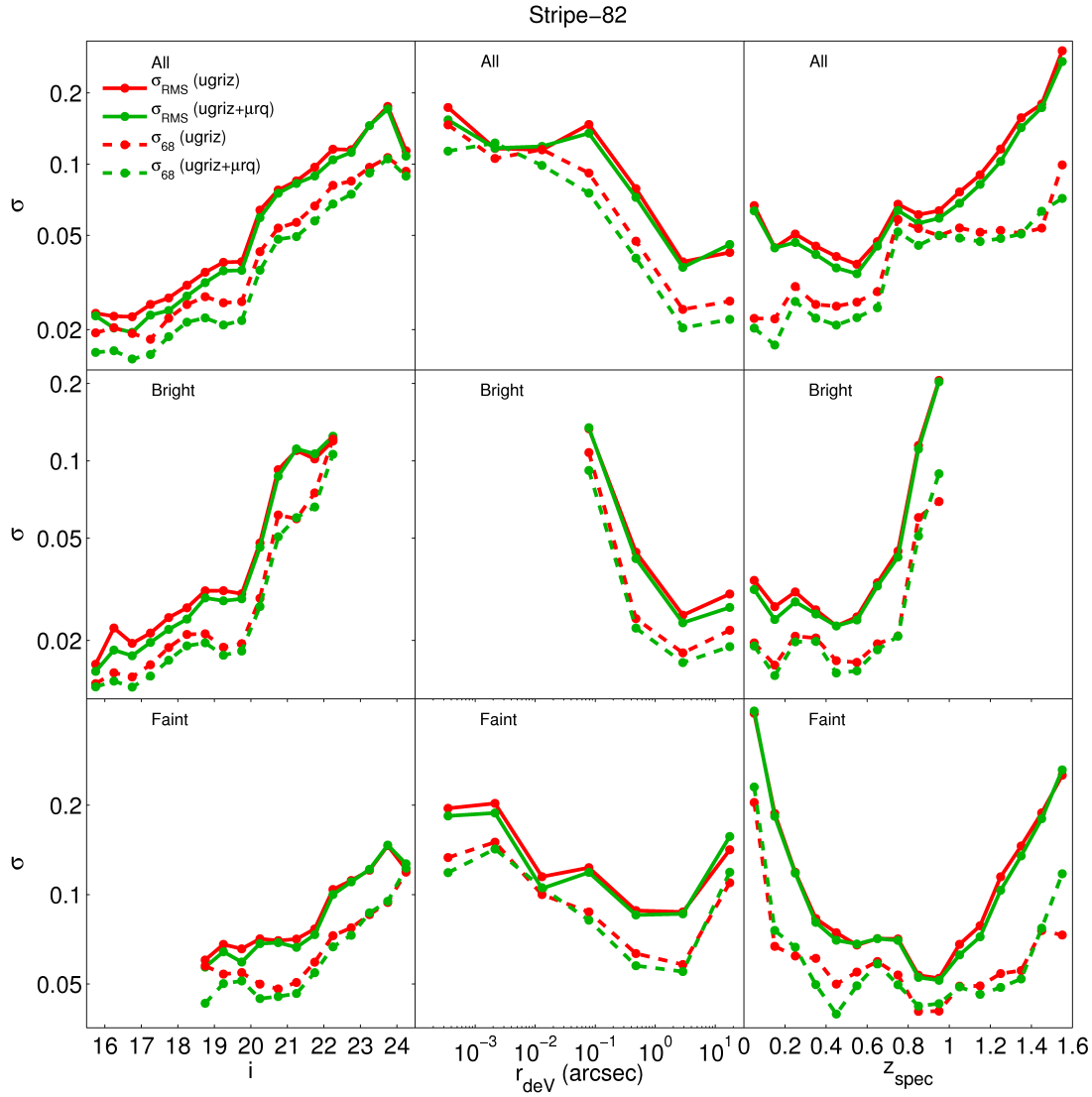

Figure 35: Change in σ_{RMS} (solid lines) and σ_{68} (dashed lines) when trained with and without morphology as it spans across the *i*-band magnitude (left), radius (middle) and redshift (right), shown for the results of the full Stripe-82 Sample (top) and its bright (middle) and faint subsamples (bottom), respectively.

We conclude that for a sample reaching $i \sim 24$, individual morphological parameters have negligible impact on photo-z's when evaluating general metrics as the improvement brought by these individual parameters are of the same order as the uncertainties of the photo-z's produced. However, the improvement is more significant when multiple morphological parameters are used, and the improvements focus on objects which are brighter, very small or very large, and have lower redshifts, which in fact are the concentration points of the bulk of the training objects. We will test this dependency of training objects in Section 6.5.1 on a similar sample but with the training set reweighted to a target sample more representative of current photometric surveys.

6.4.2 *Photometric Redshifts with Imperfect Star-Galaxy Separator*

Star-galaxy separation remains an ongoing problem when producing photo-z catalogues (Bundy et al. 2015). Current tools to separate point sources (stars and quasars) from extended objects (galaxies) include morphometric approaches, machine learning or using infrared colours (see Soumagnac et al. 2015, for a comprehensive discussion). However, since point-source separation algorithms are not perfect, the photometric galaxy catalogues produced will still be contaminated by a small number of stars or quasars, and the photo-z estimates that go with them can be wrong if the training sample is not representative of the quasar population.

To study the impact of suboptimal point-source separation, the SDSS official source classification provides a good case study. SDSS uses the field `TYPE` to indicate if an object is an extended object (`TYPE=3`) or a point source (`TYPE=6`), and thus far we have heavily relied on this star-galaxy separation flag to select our samples. Point-source contamination in a sample of `TYPE=3` objects in the SDSS Coadd data set was estimated to be as high as 10% (Bundy et al. 2015), and the impact of these objects is seen in the official SDSS Stripe-82 Coadd Photometric Redshift Catalogue (Reis et al. 2012).

For this analysis, we reproduce a similar sample to the SDSS Coadd (Reis et al. 2012), which we shall call the 'Stripe-82 Type 3 Sample'. We cross-match spec-z's of galaxies and quasars (`CLASS=GALAXY,QSO`) from SDSS, DEEP2, VVDS and WiggleZ with objects from Reis' photo-z catalogue (which are all `TYPE=3` objects). This sample selection differs from the Stripe-82 Sample used in the previous section in 2 ways: (1) spec-z which were spectroscopically identified as quasars (`CLASS=QSO`) are cross-matched as well, and (2) BOSS spectroscopic redshifts are not used here as we intend to use a spectroscopic sample as close as possible to the sample choice of Reis et al. (2012). The red line in Figure 36 shows the distributions of i and $z_{\rm spec}$ of the Type 3 Sample, noticeable is the missing 'BOSS bump' in the right panel (objects with $0.4 < z_{\rm spec} < 0.6$) when compared to the previous sample. This sample yields 97 812 objects in which only 97% have galaxy spectra. We will explain the Type 6 Sample shown in the figure at the end of this section.

Using this sample, we plot the spec-z v.s. Reis' photo-z, as shown in the top right panel of Figure 37. The blue dots in the figure evidently show that these quasars which have been photometrically identified as galaxies have their photo-z's severely underestimated, becoming outliers in our sample. If the ANN does not have high-redshift quasars for training, it will not be able to produce correct redshifts for these objects. Therefore we study the possibility of mitigating this problem by including quasar redshifts in the training, and whether size information brings

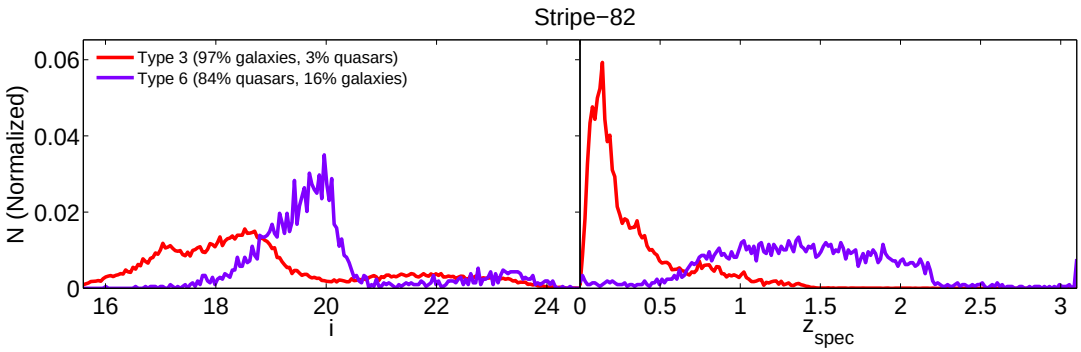

Figure 36: Distributions of the i-band (left) and $z_{\rm spec}$ (right) for the Stripe-82 Type 3 (red) and Type 6 (purple) samples, respectively.

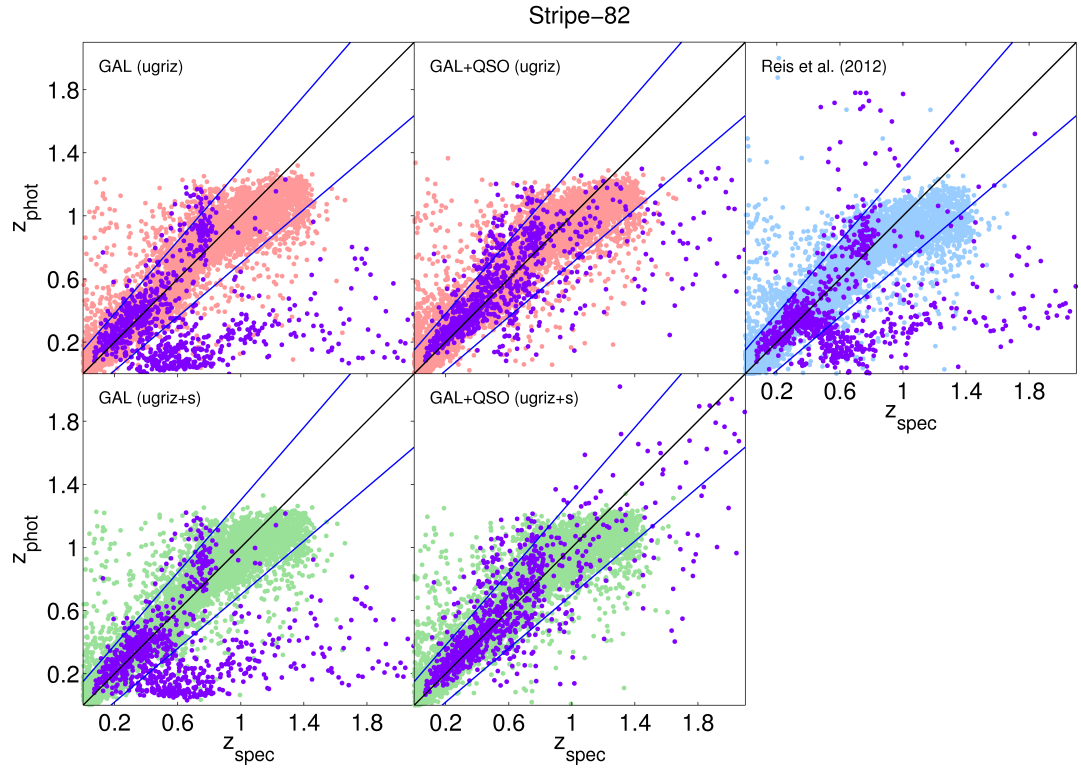

Figure 37: Photo-z v.s. spec-z of the Stripe-82 Type 3 Sample, comparing the results of ANNZ training with only galaxy redshifts (left), ANNZ training with galaxy and quasar redshifts (middle), and the results of Reis et al. (top right). The different rows show the training of *ugriz* (top), and *ugriz* with size (bottom). The inclusion of quasar redshifts (middle) has corrected photo-z's for most of the quasars (purple) that were originally on the bottom right corner of the plots. However, the joint inclusion of quasar redshifts and size in training further corrected the photo-z's for quasars with redshift $z > 1.3$.

Table 16: Improvement in $\sigma_{\rm RMS}$, σ_{68} and $\eta_{\rm out}$ when quasars or size are added to the training for the Stripe-82 Type 3 Sample.

Training objects	**Inputs**	$\sigma_{\rm RMS}$	$\Delta\%$	σ_{68}	$\Delta\%$	$\eta_{\rm out}(\%)$	$\Delta\%$
Reis et al. (2012)	$ugriz$	0.0683		0.0252		3.32	
Galaxies	$ugriz$	0.0632		0.0246		3.01	
Galaxies	$ugriz + s$	0.0622	1.6	0.0244	0.6	2.81	6.5
Galaxies + QSOs	$ugriz$	0.0598		0.0261		2.70	
Galaxies + QSOs	$ugriz + s$	0.0558	6.7	0.0259	0.8	2.40	11.1

additional improvements. We note here that the 'size' used in this sample is the circularised radius of a PSF-corrected exponential profile fit, which yield very small values for these photometrically misclassified quasars (radius derived from the bright bulge), although in some cases it is plausible that light from the host galaxy is picked up as well.

Using ANNZ, we compare 4 runs: the first two runs are trained only with galaxies, with and without size as inputs (GAL $ugriz$ and GAL $ugriz + s$); the third and fourth runs include quasar redshifts in the training, with and without size (GAL+QSO $ugriz$ and GAL+QSO $ugriz + s$). The results of these runs are summarised in Table 16 and visualised in Figure 37. We see that the improvement with size is only marginal if only galaxies are used for training. When quasars are added to the training, we correct the redshifts measured for quasars; however, this inclusion degrades the quality of all photo-z's, increasing the σ_{68} by approximately 3%. Our new important result is that the photo-z performance is recovered when we add size to the inputs: this improved the quality of the photo-z especially for quasars with $z_{\rm phot} > 1.3$.

We also find that with an 85% completeness cut in photo-z error, approximately half of the quasars can be removed from the sample, and more high-redshift objects are kept when compared to Reis et al.'s results (Figure 38). With the use of quasar redshifts and size, the outlier fraction is reduced from 3.0% to 2.4%, which is a relative improvement of 20%. This result is a clear indication that the inclusion of quasar redshifts in training may provide a more reliable photo-z for a catalogue of galaxy-like objects.

While the inclusion of quasar redshifts in training improves the overall metrics for a sample of extended objects, we are also interested to know if this has particularly degraded the photo-z quality of galaxies in the sample. We find that the degradation in galaxy photo-z's when quasars are included in training is less than 1% across magnitude and colour, which we deem insignificant. In fact, we see that the training of GAL+QSO $ugriz + s$ performs better than GAL $ugriz$ particularly for redder and larger galaxies. Galaxies which are large and red should generally have lower redshift, therefore size information helps to lower the overestimated photo-z's for these objects. More surprisingly, we also find that the training of GAL+QSO $ugriz + s$ yields better photo-z's for the `TYPE=3` quasars compared to training QSO alone.

Therefore for machine learning methods, we highly recommend the inclusion of quasar redshifts and morphology in training when estimating photo-z's for galaxy samples. Though we note that as the distribution of the training and testing set used here is entirely driven by the matching of photometry with spectroscopic samples without reweighting, we expect the results to be slightly different in a purely flux-limited sample.

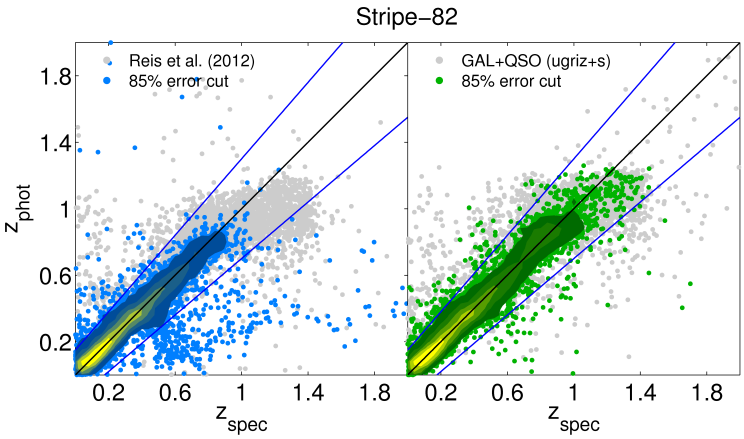

Figure 38: Photo-z v.s. spec-z of the Stripe-82 Type 3 Sample, comparing the results of Reis (left) and our results when trained with *ugriz* and size using ANNz with galaxies and quasar spectra (right). The blue and green dots show the subsamples of each when 85% of the objects with the lowest errors were kept in the sample.

Table 17: Improvement in σ_{RMS}, σ_{68} and η_{out} when galaxies or size are added to the training for the Stripe-82 Type 6 Sample.

Training objects	**Input**	σ_{RMS}	$\Delta\%$	σ_{68}	$\Delta\%$	$\eta_{\mathrm{out}}(\%)$	$\Delta\%$
QSOs	*ugriz*	0.3356		0.0837		21.64	
QSOs	$ugriz + s$	0.3306	1.5	0.0843	-0.7	18.09	16.4
QSOs + Galaxies	*ugriz*	0.1866		0.0784		16.41	
QSOs + Galaxies	$ugriz + s$	0.1782	4.5	0.0751	4.2	15.21	7.3

Note that in this study the point-source contamination problem is only tackled partially, since point-sources make up of not only quasars, but also stars. This means that for a galaxy sample contaminated by both quasars and stars, this method will still leave the stars having catastrophic photo-z's. We tried to do a similar experiment for stars as we did for quasars: we include both `TYPE=3` stars and quasars in the training (stars have $z_{\mathrm{spec}} \approx 0$ and extremely small sizes), and see if the ANN was able to identify the stars give them $z_{\mathrm{phot}} = 0$. However, the experiment did not turn out well: the overall scatter and outlier fraction rate have degraded, and the photo-z's of these stars were not estimated correctly. Therefore, we note that the method shown in this section will only work correctly with the assumption that stars have been separated from both the galaxies and the quasars in the sample prior to training.

As a bonus to this section, we also tried the converse: we wanted to see if morphology improves the photo-z's of photometrically identified quasars (point sources) when using quasar and galaxy spectra. This is done by selecting a spectroscopic sample with `TYPE=6` and spectra with `CLASS=GALAXY,QSO` using the same sources of photometry and spectroscopy as before, which we call the Stripe-82 Type 6 Sample. This yields a small sample of 6 252 objects (distributions of i and z_{spec} shown as the blue line in Figure 36). This sample is contaminated by 16% galaxies. Once again we used ANNz to train these objects with and without galaxy spectra, with and without size, and the results are shown in Table 17 and illustrated in Figure 39.

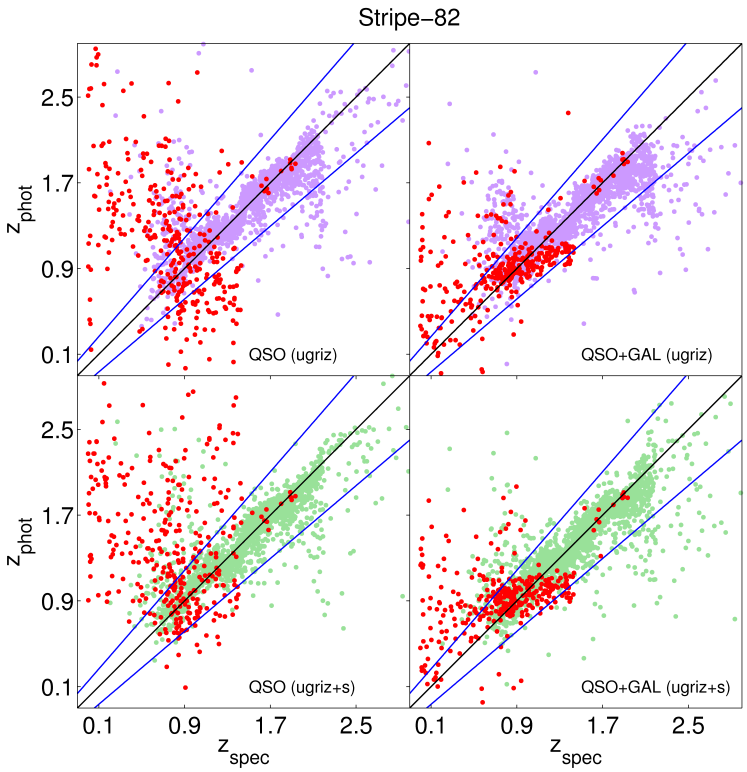

Figure 39: Photo-z v.s. spec-z of the Stripe-82 Type 6 Sample, when training includes quasars (left) or a mixture of quasars and galaxies (right), comparing the results when trained with *ugriz* (top) and *ugriz* with size (bottom). The red dots are objects with galaxy spectra.

The results seem to be similar to the case of the Type 3 Sample: while size does not improve the quasar photo-z's significantly, a combination of galaxy and quasar spectra trained with only *ugriz* showed significantly improved results. When size is included in training in the case of training with mixed galaxies and quasars, we see also see improvement to the photo-z, particularly the scatter of quasars in the mid-redshift range is reduced. We find this results interesting and believe that it may be beneficial to the quasar community.

Therefore we conclude this section by stating that size brings about more improvement to photo-z's in galaxy samples when used concurrently with quasar spectra, especially when the star-galaxy separation is imperfect. We have applied this technique to train and obtain photo-z's for the CS82 Morphology Catalogue, which will be discussed in Section 7.3.

6.4.3 *Photometric Redshifts with Infrared Magnitudes*

It has been shown inBrammer et al. (2008) and Bundy et al. (2015) that the inclusion of infrared photometry like the bands $YJHK$ have degraded photo-z's for template-based codes, and this is mainly due to template inaccuracies. However, we assert that this may be different for the case of machine learning photo-z's, particularly if the optical and infrared photometry from two different surveys are calibrated and well matched. Thus far we have been working with SDSS photometry, which consists of only *ugriz* broadband magnitudes. In this section we attempt to conduct a short test on the performance of photo-z's when infrared photometry and galaxy size are added concurrently in training.

Continuing on our focus on the Stripe-82 region, we utilise photometry provided by the S82-MGC (Bundy et al. 2015). This catalogue provides photometry from the SDSS Coadd and infrared photometry from UKIDSS, these photometry are matched using Synmag (Bundy et al. 2012), a synthetic aperture photometric matching technique which works at the catalogue level without requiring reprocessing of the images. These reprocessed and matched synthetic magnitudes *ugrizYJHK* provide an excellent dataset for the purpose of this analysis. We cross-match these magnitudes with SDSS galaxy sizes and galaxy spectroscopic redshifts from SDSS, BOSS, DEEP2, WiggleZ and VVDS to form a sample of 62 603 objects, divided equally for training, validating and testing. The distributions of i and $z_{\rm spec}$ of the S82-MGC Sample used in this work is shown in Figure 40. It is worth noting that this sample is relatively brighter than the Stripe-82 Sample used in Section 6.4.1, this is due to the cross-matching selection effects between SDSS and UKIDSS photometry.

We used Annz to estimate photo-z's for the sample when trained using *ugriz* and *ugrizYJHK*, with and without size, and the results are shown in Table 18. Overall we see that with the inclusion of infrared bands, the photo-z see great improvement, and when size is added to the training, the results get even better, reaching a σ_{68} of 0.0158. We compare our results with the photo-z's provided by the S82-MGC, through the template-based codes Eazy, Bpz and photo-z's from the red-sequence Matched filter Galaxy Catalogue (redMaGiC, Rozo et al. 2016). From Figure 41 we see that our photo-z's produced using *ugriz* looked comparable to the results of Bpz, the best among them all, however, when infrared magnitudes and size are included in the training (green plot), we see the objects moved closer to the diagonal. We note here that the

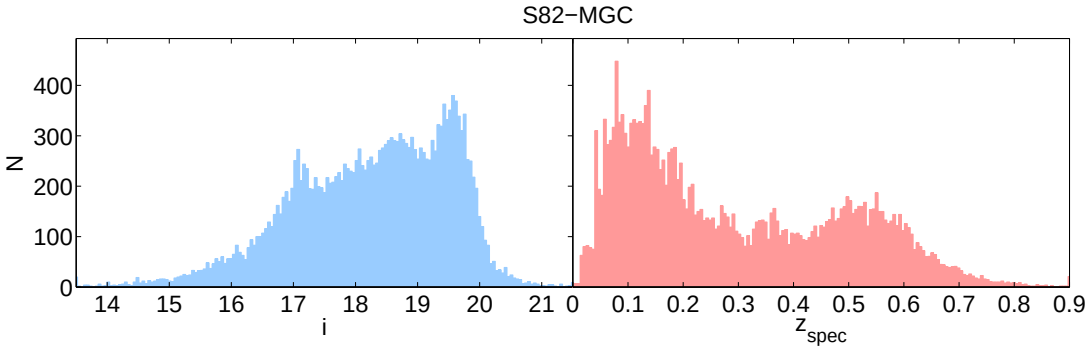

Figure 40: Distributions of the i-band (left) and $z_{\rm spec}$ (right) for the S82-MGC Sample.

Table 18: Relative improvement in $\sigma_{\rm RMS}$, σ_{68} and $\eta_{\rm out}$ for the S82-MGC Sample when trained with and without infrared photometry, with and without size. Also shown are the metrics for photo-z's by EAZY, redMaGiC and BPZ as provided in the S82-MGC.

Input	$\sigma_{\rm RMS}$	Δ%	σ_{68}	Δ%	$\eta_{\rm out}(\%)$	Δ%
$ugriz$	0.0290		0.0204		0.34	
$ugriz+s$	0.0275	5.0	0.0179	12.3	0.35	-1.4
$ugrizYJHK$	0.0245		0.0161		0.27	
$ugrizYJHK+s$	0.0236	3.4	0.0158	2.0	0.23	15.8
EAZY	0.0519		0.0318		0.97	
redMaGiC	0.1559		0.0523		6.27	
BPZ	0.0471		0.0314		0.80	

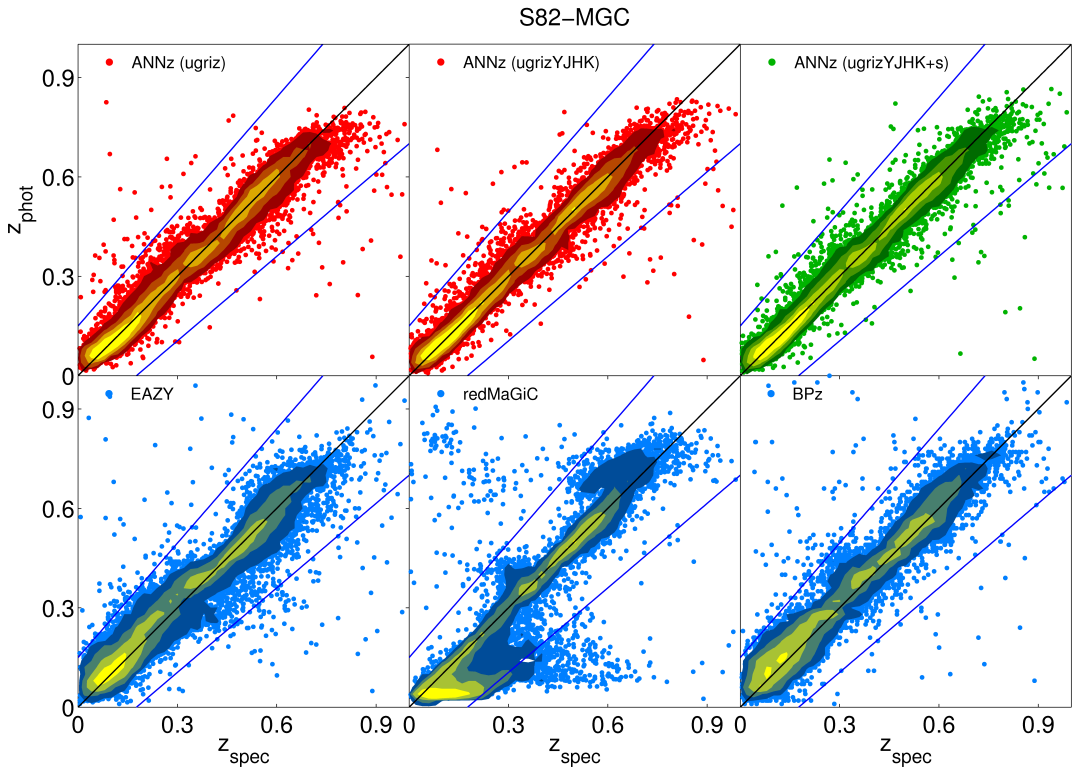

Figure 41: Photo-z v.s. spec-z of the S82-MGC, comparing training with *ugriz* (top left), *ugrizYJHK* (top middle) and *ugrizYJHK* with size (top right). Also shown in the bottom row are photo-z's provided by the catalogue, showing the results of Eazy (bottom left), Bpz (bottom right), and redMaGiC (bottom middle).

photo-z's produced by Eazy and Bpz for the catalogue did not use infrared photometry as they did not improve the results (Bundy et al. 2015).

We conclude from this section that morphology improves photo-z's on top of the improvement infrared photometry brings, and the results of Annz remain competitive when compared to template-based codes. Thus far we have established that galaxy morphology always brings about positive impact on photo-z's of any sample with $i < 20$, we therefore recommend this to be common practice for all machine learning photo-z estimations, provided that representative training sets are available. Since galaxy morphology is usually measured in conjunction with or as a by-product of photometry, it would definitely be a waste if they were not put to good use.

6.5 THE CS82 SAMPLE

Up until this point of the chapter, all of the spectroscopic samples we have used in our analysis are based on samples which have spectroscopic redshifts in abundance, or surveys with a relatively bright magnitude limit by current standards. In order to keep the relevance of this analysis with current and future surveys where observations go deeper and fainter and where spectroscopic training objects are scarce, we are interested to study the impact of galaxy morphology on samples which would be representative of modern surveys.

As we have seen in the previous section, spectroscopic galaxy samples are usually constructed by cross-matching photometry from large photometric surveys with redshifts obtained from multiple different spectroscopic surveys. Due to the different target selections of each spectroscopic survey, the combined spectroscopic training sample will contain objects which are unevenly distributed in colour-magnitude space, e.g. contain preferentially bright and red galaxies. This means that the distributions of training parameters in magnitude, colour, size, etc. will turn out to be quite different from that of the target sample, in which the photo-z's are to be estimated. Therefore, reweighting of the spectroscopic sample to become representative of the target sample is needed, not only to ensure that the metrics evaluated on the testing set is representative of the target set, but also to ensure that none of the spectroscopic sources are over-represented in the training.

For the purpose of this analysis we have selected the CS82 Survey (Moraes et al. 2014) to be our target set where we would want our training and testing to be representative of. The CS82 Survey is a single i-band survey, it has a seeing as low as $0.6''$, it produced high quality morphology, and its observations reach as deep as $i \sim 24.1$ (see Section 4.1.4 for more details). We have selected a sample of galaxies close to representing current large-scale galaxy surveys like KiDS and DES, although we note that this sample does not cover the range of magnitudes expected from Stage IV surveys such as LSST.

To construct a spectroscopic sample for this analysis we use *ugriz* photometry from SDSS Stripe-82 Coadd, morphology from the CS82 Morphology Catalogue (with cuts described in Section 4.1.4), and galaxy spectroscopic redshifts from SDSS, BOSS, DEEP2, WiggleZ and VVDS. We further make a magnitude cut of $16.0 < r < 24.5$ to remove bright objects and unreliable faint data. We make a special note here that for this sample we do not use the star-galaxy separation tag of SDSS, but instead use `SPREAD_MODEL_SER>0.008` from the CS82 Survey, where `SPREAD_MODEL_SER` is the Sérsic spread model, a star-galaxy separator to select only extended objects, with its justification discussed in Moraes et al. (in prep.).

Besides, the exponential radius and axis-ratio from CS82 are used instead of their de Vaucouleurs counterparts as they are found to perform better for this sample. Small selection effects may arise since objects with bad galaxy morphology data from CS82 have to be removed. The selection cuts above produce a sample of 59 498 galaxies, equally divided into training, validating and testing to be used in the following sections. For comparison, this sample differs from the Stripe-82 Sample used earlier in the presence of CS82 morphology, an extra magnitude cut, the use of exponential fit morphology and a different star-galaxy separator.

In this section we adopt the reweighting method as introduced by Lima et al. (2008), which is done by comparing the density of objects in a selected parameter space of the spectroscopic and target samples, and setting a weight value to each object in the spectroscopic sample so that during the training process, the cost function used to estimate the photo-z will be balanced by upweighting objects that are less represented in the training sample compared to the target sample, and downweighting objects otherwise (see Section 3.3.4 for more details). Since the testing set (where performance metrics are evaluated) is also drawn from the spectroscopic sample, the weights have to be taken into account when estimating the metrics to reflect the photo-z performance of the target sample (see Sánchez et al. 2014, for more details).

From this section onwards we switch from ANNZ to ANNZ2 as it allows weights to be incorporated during the training process. However, the in-built reweighting code in ANNZ2 is not used

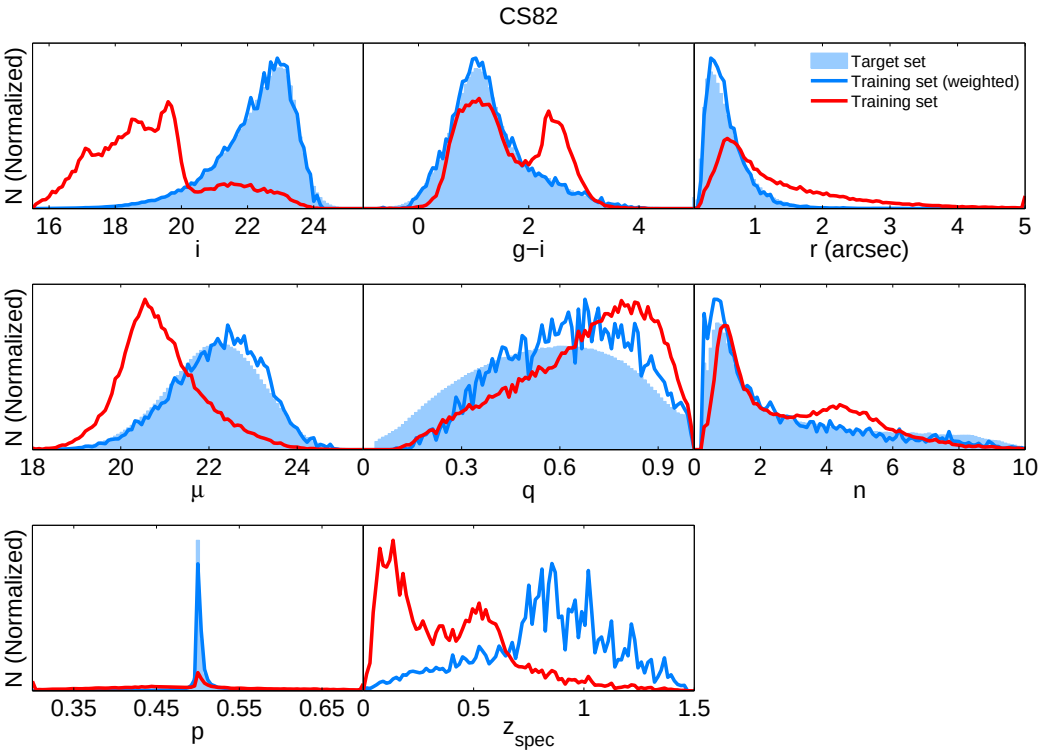

Figure 42: Distributions of i-band magnitude, colour g-i, r_{exp}, μ, q, n, p and z_{spec} for the target set (blue histogram), compared to the distributions of the training set, both weighted (blue line) and unweighted (red line). Note the reweighting is only done in terms of i and g-i, but works well for all other parameters considered.

in this project, we use an external reweighting algorithm similar to one used in Sánchez et al. (2014) to calculate the individual weights of the objects. This algorithm first uses a k-dimensional tree to bin the objects in the parameter space assigned. It then proceeds to calculate the number of nearest neighbours of each object in the sample. The weights are then derived by calculating the ratio of the densities between the training and the target sample. The weights obtained are then used in ANNZ2 to calculate the photo-z's.

We construct weights for each training and testing object by reweighting them with respect to the CS82 Sample in only the i-band and g-i colour, as we have tested and find that reweighting in more than the 2 parameters mentioned do not show any significant difference. It was also found that the other input parameters are well reweighted by just using these two parameters alone, as seen in Figure 42, where we also notice the huge difference between the distribution of i for the training set and the target set. From this figure we confirm that the reweighting in just i and g-i is able to reweight all other morphological parameters to represent the distribution of the target set. In the final panel we also see the expected redshift distribution of the target set, which should peak around $z_{spec} \approx 0.8$. We also refer the reader to Figure 43, which shows the correlation between these morphological parameters and spectroscopic redshift, both in weighted and unweighted densities. From the plot we can see that the correlation between most of these parameters and z_{spec} have weakened when weighted.

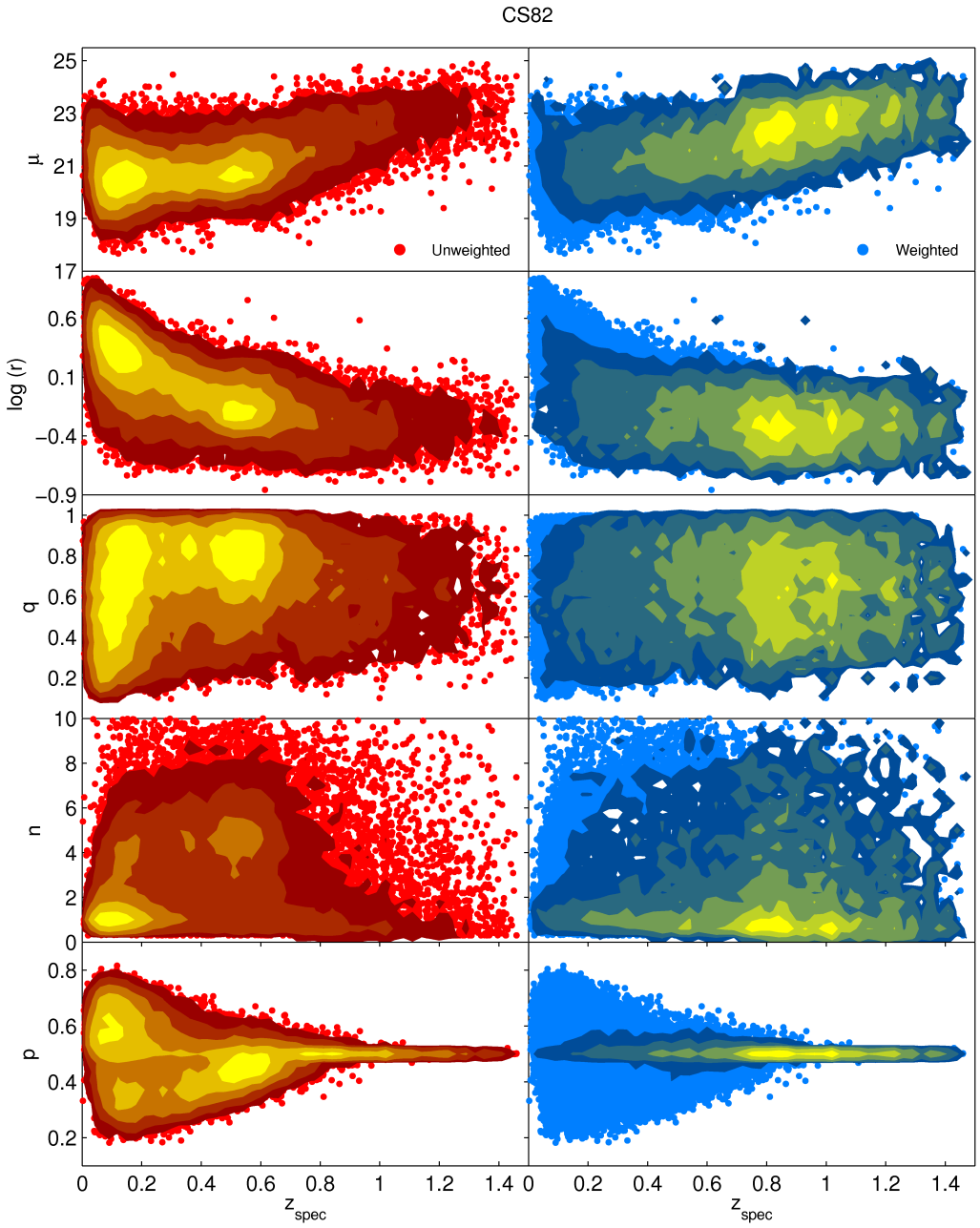

Figure 43: Scatter plots of log-radius $\log_{10}(r)$, μ, q, n and p against $z_{\rm spec}$ for objects in the CS82 photo-z training set. The contours represent the unweighted (red, left) and weighted (blue, right) density of objects, weighted with respect to the CS82 Target Sample.

Table 19: Improvement through morphology information in $\sigma_{\rm RMS}$, σ_{68} and $\eta_{\rm out}$ for the weighted CS82 Sample, with respect to training with only *ugriz*.

Inputs (*ugriz*+)	$\sigma_{\rm RMS}$	$\Delta\%$	σ_{68}	$\Delta\%$	$\eta_{\rm out}(\%)$	$\Delta\%$
-	0.0921		0.0625		6.42	
μ	0.0939	−2.0	0.0616	1.4	6.67	−3.9
r	0.0933	−1.3	0.0635	−1.1	6.18	3.8
$s_{\rm SDSS}$	0.0925	−0.5	0.0609	2.5	5.95	7.4
$s_{\rm CS82}$	0.0924	−0.3	0.0611	2.3	6.31	1.7
q	0.0940	−2.1	0.0629	−0.6	6.33	1.5
n	0.0928	−0.8	0.0626	−0.2	6.31	1.7
p	0.0946	−2.7	0.0625	0.1	6.39	0.4
r, q	0.0940	−2.1	0.0595	4.7	6.72	−4.6
r, q, μ, n, p	0.0914	0.7	0.0604	3.4	6.15	4.2

6.5.1 *Improvement by Morphological Parameter with Reweighting*

The first question we would like to address is whether the addition of morphological quantities to neural network training helps to obtain better redshifts in a weighted sample. We perform several runs with input parameters $ugriz + m$, where m is a single (or a set of) morphological parameter(s) from CS82, with the exception of size, where we also use the size from SDSS for comparison. These results are then compared to training with *ugriz* inputs alone.

Table 19 shows the metrics $\sigma_{\rm RMS}$, σ_{68} and $\eta_{\rm out}$ of the different photo-z trainings for comparison. We remind the reader that the training in this case has been reweighted such that the cost function of the ANN would focus on small changes in high-weighted objects, and the metrics shown here are also reweighted to represent the performance as if tested on the target set. In contrast to previous analyses, we see no significant improvement overall, and this is true even for surface brightness and size which were shown to bring the most improvement in the previous sections. When adding the whole set of morphological parameters chosen for this study, we reach only about 4% improvement in $\sigma_{\rm RMS}$ and 3% improvement in σ_{68}. Here we have also tested our runs with different versions of size estimators, including radii from different model choices and circularised radii from both SDSS ($s_{\rm SDSS}$) and CS82 ($s_{\rm CS82}$). Although the CS82 Survey's average seeing is half that of SDSS, this does not have significant impact on our results.

As we were interested to compare weighted results with that of an unweighted one, we reran the results for the case of *ugriz* and $ugriz + \mu rqnp$ without any reweighting with ANNZ2, and we illustrate the results in Figure 44. We noticed that the improvement for the unweighted case was small as well, yielding an improvement in $\sigma_{\rm RMS}$ of about 3%. However from the plot we see the improvement for the weighted case is in fact visible, when trained with morphology, the densities are more aligned to the diagonal line for the redshift range $0.7 < z_{\rm spec} < 1.3$, though mostly dominated by noise.

We interpret our results as stating that training with just photometry saturates the available redshift information for a galaxy population typical of Stage-II and Stage-III optical galaxy surveys, and that morphology does not significantly help to improve photometric redshift estimation beyond this. In past investigations, it was clear that most improvements brought by

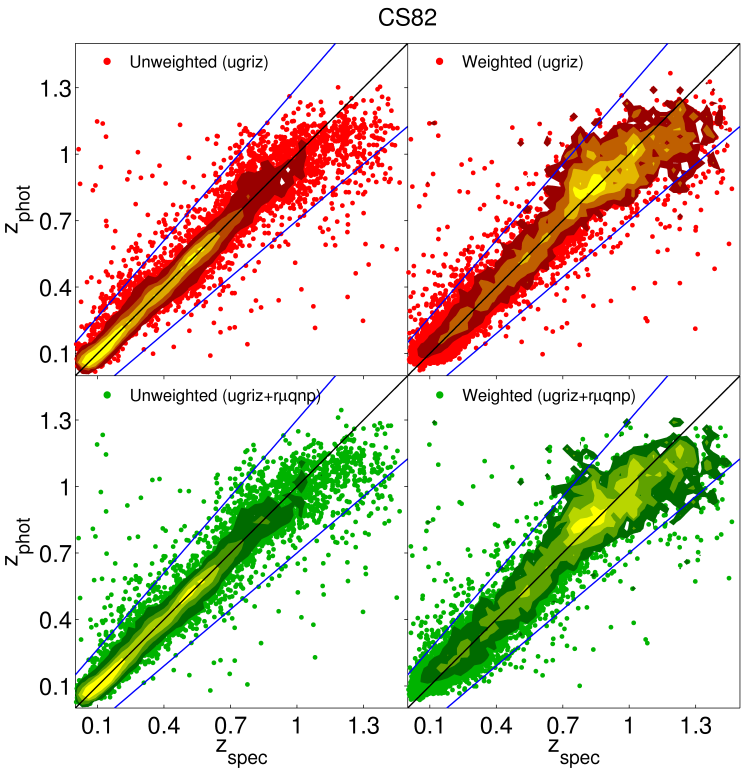

Figure 44: Photo-z v.s. spec-z of both the unweighted (left) and weighted (right) CS82 Sample, comparing the results when trained with *ugriz* (top) and *ugriz* with morphology (bottom). The coloured contours for the weighted case are weighted densities with respect to the CS82 Target Sample.

morphology were seen for bright SDSS samples only. Therefore selection through cuts in flux, morphological parameters and indirectly through the selections of spectroscopic samples will have a strong impact on the outcome. We will however show in the next 2 sections that, as the availability or quality of flux information degrades, adding morphological quantities brings quantitative and qualitative improvements to the redshift estimation process. We explore the possibilities of using morphology to improve photo-z's in suboptimal conditions.

6.5.2 *Photometric Redshifts with Limited Number of Filters*

In the previous section, we have seen that galaxy morphology has only marginal impact on photo-z quality when tested in a weighted sample of galaxies with good 5-band *ugriz* photometry. It is generally accepted that 4 or – ideally – 5 photometric bands are necessary for measuring photometric redshifts with the accuracy required by the main scientific goals of modern galaxy surveys. For instance, weak lensing surveys like DES, LSST or KiDS for which coarse line-of-sight resolution is sufficient require coverage from near-ultraviolet to near-infrared in at least 4 to 5 bands (Abbott et al. 2005; Ivezić et al. 2008; de Jong et al. 2013). This has been empirically supported in several analyses of fewer-band surveys and in extensive studies of photo-z robustness under different observational conditions, thus informing the design of some of the key experiments of the coming decade.

There are, nonetheless, design choices or technical issues that might constrain surveys to work with fewer bands than would be optimal. DECaLS (Schlegel et al. 2015) and the Canada-France Imaging Survey (CFIS, Ibata et al. 2017) are examples with limited filter coverage. Technical issues can also prevent the full exploitation of survey data, such as the limited depth of SDSS z-band due to filter and CCD inefficiencies, or the incomplete coverage in r and i-bands of the RCS-2 (Gilbank et al. 2011) due to bad seeing.

Within this context, we ask whether the addition of morphology in few-band scenarios can mitigate the degradation due to the lack of detailed colour information. We perform several ANNZ2 runs with different combinations of a smaller number of bands, both with and without morphology, and compare the overall performances of these runs. For morphology, we use all 5 morphological parameters (μ, r, q, p and n) together. Figure 45 shows the performance metrics of these runs with respect to the choice of filters used. All metrics improve with morphology relative to the photometry-only case, and more so as the number of broadband filters decreases. Taking the case where only *grz* bands are available (similar to the case of DECaLS), we see about 14% improvement in σ_{68} and 18% improvement in outlier fraction when morphology is included in training, reaching a performance in all 3 metrics that is close to the full *ugriz* case without morphology. Furthermore, a training with 1 colour (*ri*) and morphology performs at least as well as training with 3 filters and no morphology. In the face of these results, there is a strong case for using morphology in photo-z estimation in surveys which have limited multi-band photometry, like the Red Cluster Sequence Lensing Survey (RCSLenS, Hildebrandt et al. 2016), DECaLS and the Beijing-Arizona Sky Survey (BASS, Zou et al. 2017).

We explore in more detail some specific qualitative aspects of particular relevance. Figure 46 compares the results with 5 (*ugriz*), 3 (*grz*) and 1 (*i*) band(s). With *grz*, we see that the addition of morphology has a very noticeable impact at higher redshifts. Possibly due to biases deriving

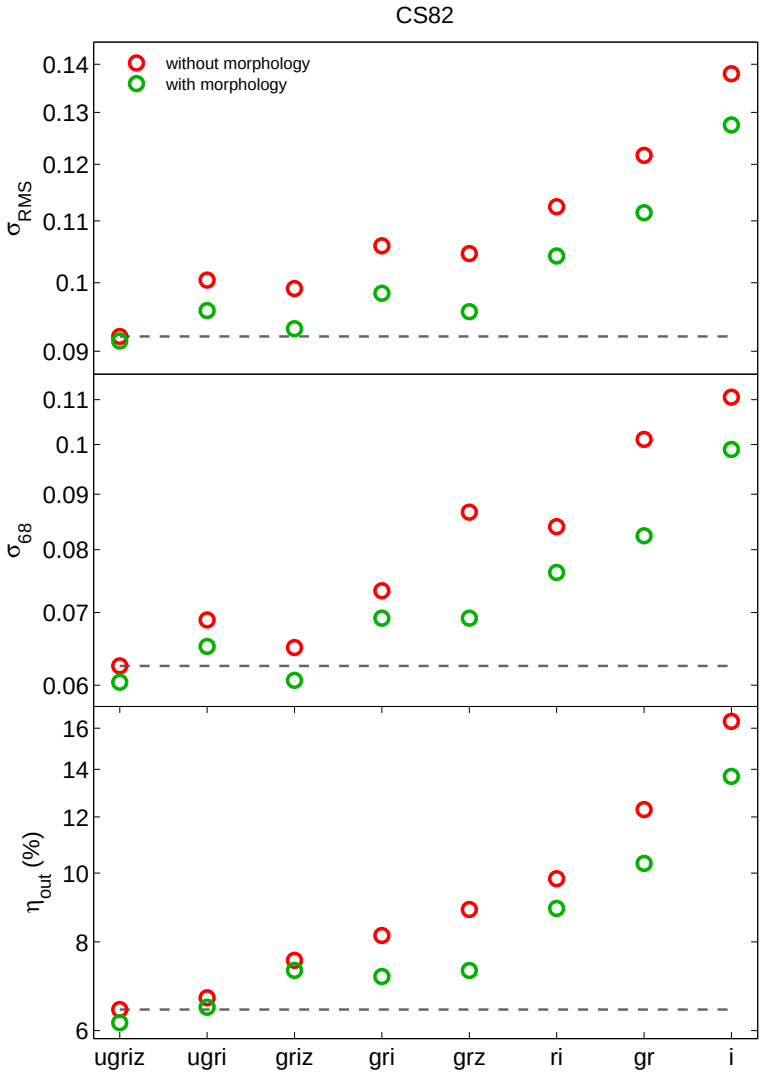

Figure 45: The comparison of $\sigma_{\rm RMS}$ (top), σ_{68} (middle) and $\eta_{\rm out}$ (bottom) for different photo-z runs. Each panel compares runs with pure photometry (red) to runs with colour and morphology (green) for each combination of bands. The pure *ugriz* run is also shown as a horizontal grey dashed line.

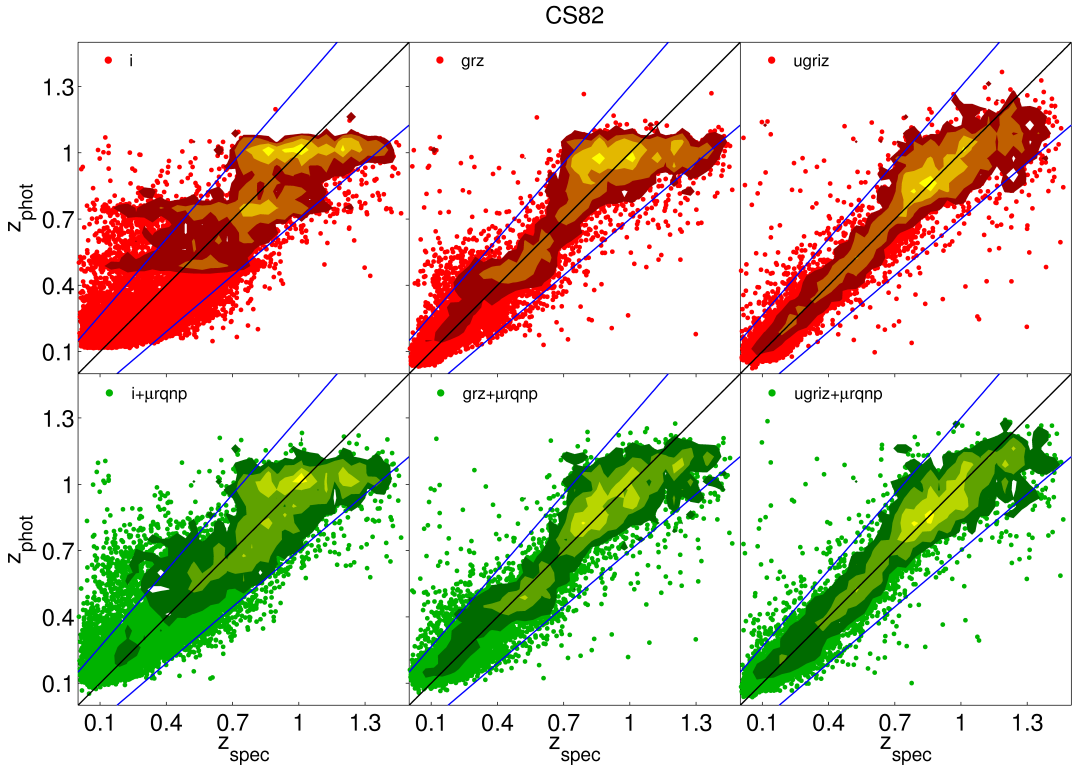

Figure 46: Comparison between photo-z and spec-z based on the i (left), grz (middle) and $ugriz$ (right) band(s). The top panels show the training with only the respective magnitude bands, while the second row shows the training when the five morphological parameters (r, q, μ, n and p) are included.

from the shallowness of the z-band in the SDSS Survey, as shown by the contours in the figure, the ANN 'saturates' after a certain redshift value, never assigning higher values; the addition of morphology redresses this high-redshift problem. Even more striking is the case with only one band i. As Figure 46 shows, a single apparent magnitude provides no more than a coarse indicator for a galaxy's redshift. However, with the addition of morphological parameters, redshifts can be measured at a level of precision and accuracy that, although far from the best scenarios, makes them usable for defining broad redshift bins.

Overall, we see a robust trend where morphology provides complementary information to colours, such that the removal of colour information can be compensated by adding morphological information.

6.5.3 *Photometric Redshifts with Low-Quality Photometry*

The quality of photo-z's is highly dependent on the quality of the multi-band photometry. Surveys relying on ground-based observations will inevitably accumulate data in a variety of conditions, resulting in a spatially-varying fidelity and signal-to-noise of the photometry. We will explore for an illustrative case if galaxy morphology is able to salvage the quality of photo-z's in situations of poor photometry.

Once again Stripe-82 is one of the best regions of the sky to conduct this study due to the multiple repeated scans across this region. Prior to Fall 2004, all observational runs were taken under photometric conditions as required for imaging in the SDSS Legacy Survey (York et al. 2000). Repeated imaging from these 84 runs and a few later runs with seeing better than $2''$, sky brightness less than 19.5 mag arcsec^{-2} and extinction less than 0.2 mag were processed for co-addition (Annis et al. 2014), which is the photometry used throughout Section 6.4 and Section 6.5. By our standards, this coadded photometry is considered 'good photometry', it was designated `run=106,206` in the SDSS CAS, and reaches approximately 2 magnitudes fainter than the SDSS single runs, and a median seeing of $1.1''$ (compared to the usual $1.4''$).

In contrast to this good photometry, runs later than Fall 2005 on Stripe-82 were made as part of the SDSS Supernova Survey (Frieman et al. 2008), and observations were done with a higher cadence and often observed under poor seeing conditions ($\geq 2''$), bright sky, non-photometric conditions and low atmospheric transparency (Sako et al. 2008). This 'bad photometry', although having photometric errors at least twice as large as those from the co-add runs, was still used in science analyses after images taken under extremely poor photometric conditions were removed, and the remaining detections subjected to a photometric calibration procedure (Bramich et al. 2008). This resulted in a photometry subset with larger magnitude errors, especially in the u-band.

We constructed a sample with 'bad photometry' as follows: instead of matching our spectroscopy and CS82 morphology with the good photometry from the co-added runs (106 and 206), we matched them to photometric objects without restricting from which run the object's measurements were taken, but still ensuring that these were `PRIMARY` objects. This way, objects with the same spectroscopy and morphology data have been matched to two different kinds of photometry, in one case obtained from the SDSS Coadd Sample, and in the other from runs with lower quality. This allows us to compare the photo-z performance of the same objects under the

Table 20: The mean magnitude errors for each band, when compared between the good and bad photometry samples.

Filter	Mean magnitude error	
	Good photometry	Bad photometry
u	0.447	0.762
g	0.031	0.135
r	0.017	0.061
i	0.015	0.049
z	0.038	0.128

Table 21: Improvement in σ_{RMS}, σ_{68} and η_{out} by morphological parameter and number of filters for the CS82 Sample with low-quality photometry.

Inputs (*ugriz*+)	σ_{RMS}	Δ%	σ_{68}	Δ%	$\eta_{\mathrm{out}}(\%)$	Δ%
-	0.1117		0.0892		12.86	
r	0.1203	−2.2	0.0893	−0.2	14.46	−12.4
s_{SDSS}	0.1138	3.3	0.0864	3.2	12.64	1.7
s_{CS82}	0.1147	2.6	0.0868	2.7	12.85	0.1
μ	0.1137	3.4	0.0863	3.2	12.03	6.5
q	0.1163	1.2	0.0877	1.6	13.73	−6.7
n	0.1142	3.0	0.0856	4.0	13.14	−2.1
p	0.1160	1.5	0.0863	3.2	13.27	−3.2
r, q	0.1130	4.0	0.0861	3.5	12.46	3.1
r, q, μ, n, p	0.1093	7.2	0.0827	7.3	11.04	14.2

same reweighting scheme, but with different photometric quality. The magnitude errors for each band are provided in Table 20, while the magnitudes of each band for for overlapping galaxies from both the good and bad photometry samples are compared and shown in Figure 47. It can be seen that the difference in u-band magnitude is much higher than other bands, and the bad photometry sample has mean magnitude errors 2 to 3 times larger than those of the good photometry sample. This sample yields 57 784 galaxies, of which 58% of the objects have low-quality photometry. There are more objects with non-detected morphological data in this sample than the former, thus the removal of these objects resulted in a slightly smaller sample.

We conduct the exact same test as we did for the good photometry in Section 6.5.1 and Section 6.5.2, adding various different morphological parameters in training while also varying the number of broadband filters used. Table 21 shows the results of this run, which is a direct comparison with Table 19. We see a higher improvement rate in photo-z in the bad photometry sample compared to the good one, especially when all 5 morphological parameters are included, yielding a relative improvement in outlier fraction as high as 14.2%. We also see that even when low-quality morphology is used (s_{SDSS}) in this case, the improvement brought is still higher than the former, although not very significant. However, it is evident that all metrics in this sample are worse than the good photometry case, even with the help of morphology.

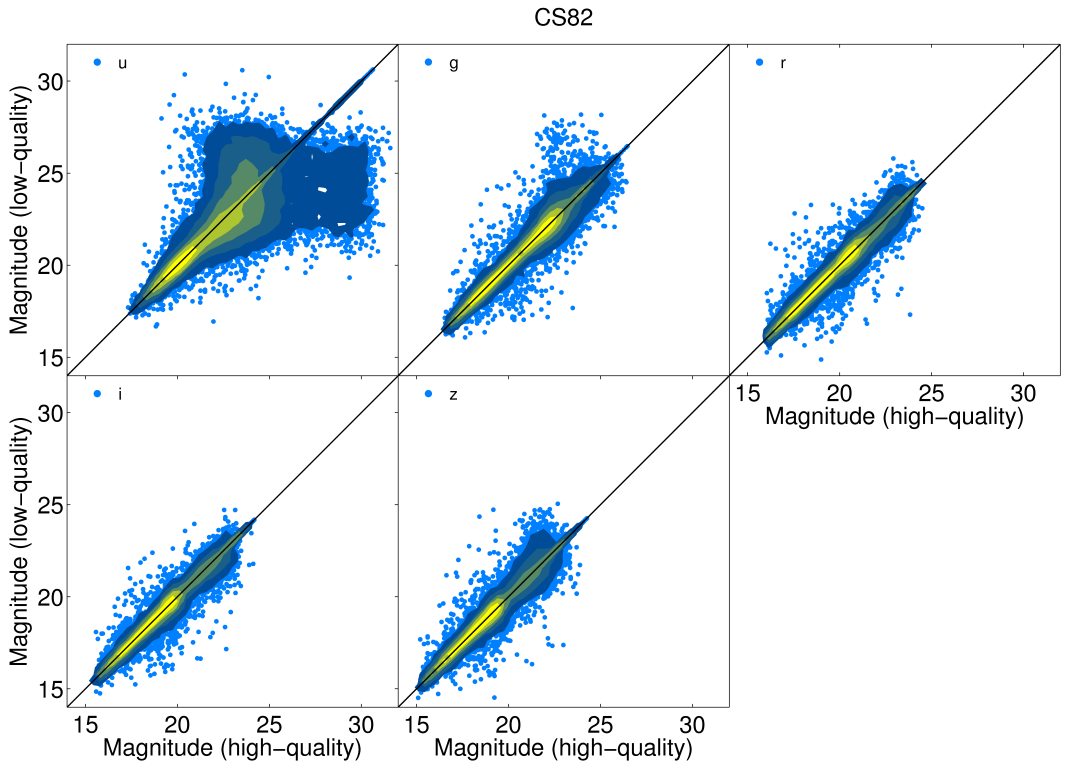

Figure 47: Comparison between the high-quality ('good') photometry (from Stripe-82 Coadd) and the low-quality ('bad') photometry (from the Supernova Survey) for overlapping objects within both samples. The difference between the u-band magnitudes is visibly large.

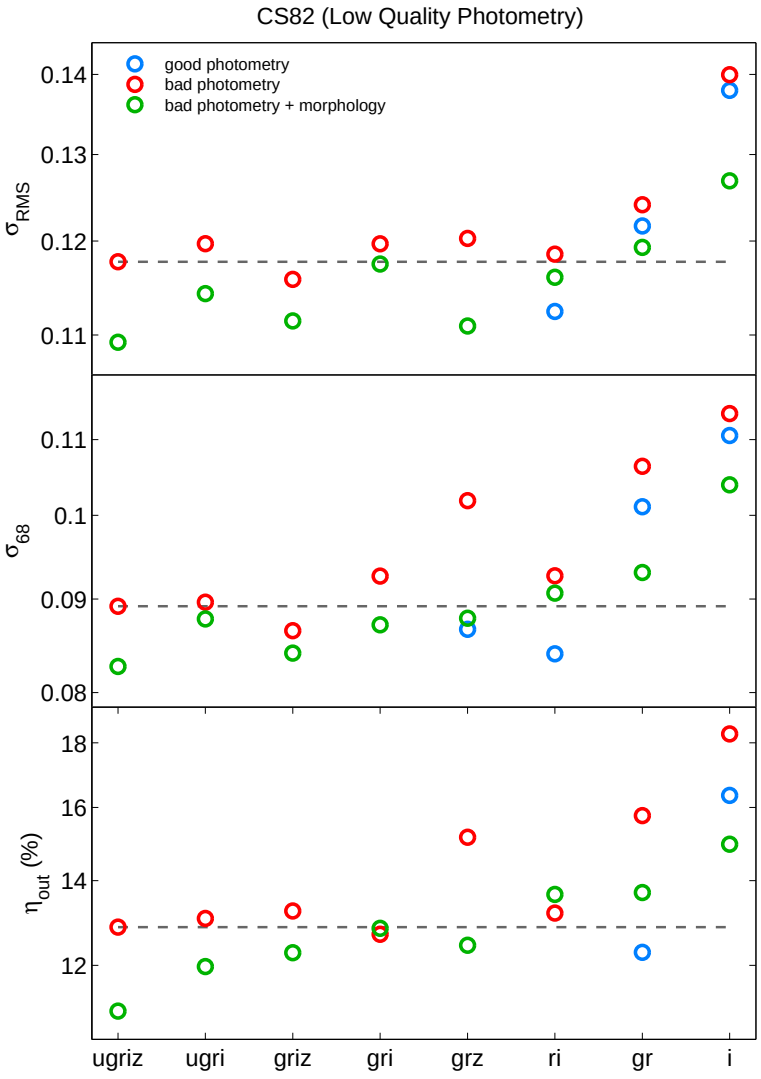

Figure 48: Comparison of σ_{RMS} (top), σ_{68} (middle) and η_{out} (bottom) for photo-z training with different number of filters. Each panel compares runs with pure photometry (red) to runs with colour and morphology (green) for each combination of bands for the low-quality photometry set, further compared to runs using pure photometry from the good quality photometry set (blue). The pure *ugriz* run for the bad photometry case is also shown as a horizontal grey dashed line. Note that some blue circles are not shown as they are far below the grey line.

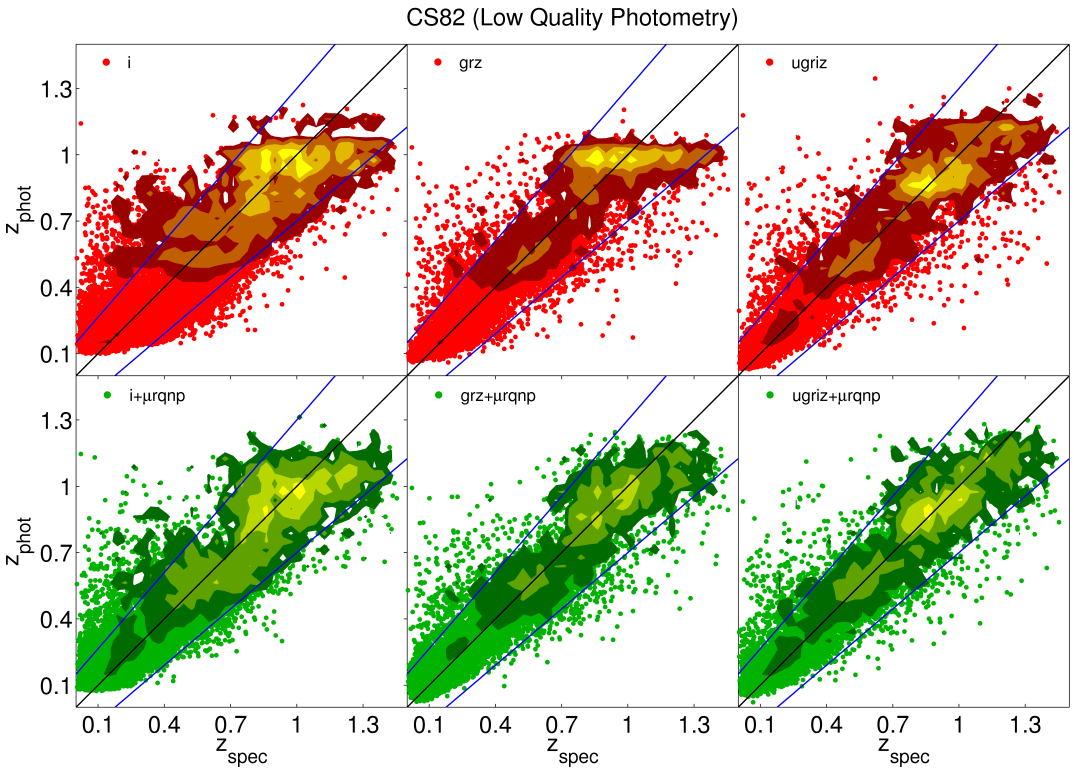

Figure 49: Photo-z v.s. spec-z of the CS82 Sample when trained with low-quality photometry, comparing the results when trained using different number of bands, namely i (left), grz (middle) and $ugriz$ (right). The results when trained with $ugriz$ (top) is compared to those when trained with $ugriz$ and morphology (bottom).

We also assess the case when fewer filters are used with and without morphology, and the results are summarised in Figure 48. Here we see that the when fewer filters are used, the improvement that morphology yields is on average larger than when good photometry is used, especially for the $ugriz$, grz and i cases. It is worth noting that in this bad photometry case, photo-z's produced with only 2 bands and morphology can yield performance metrics as good as 5 bands without morphology. We see that the metrics here are still far from the case when good photometry is used (blue circles), except for the cases when less than 3 filters are used.

These results further strengthen the case that morphology is a valuable addition to improve the quality of photo-z's under suboptimal conditions. We also show Figure 49, a figure in direct comparison to Figure 46 for the sample with good photometry, where substantial improvement is observed when morphology is added to the training of i, grz and $ugriz$.

6.6 CONCLUSION

In this chapter we have conducted a thorough analysis on the impact of galaxy morphology on various samples of different depth and redshift. We summarize the results of our 4 main sections in the points below:

1. For the MGS ($i < 17.5$, $z_{\rm spec} < 0.3$), surface brightness and size improve photo-z's the most, and when multiple morphological parameters are used, the results are even better. Morphological parameters also bring improvement to photo-z's of elliptical and spiral galaxy samples, and more improvement is found in the latter. The fundamental plane relation did not improve photo-z's in the MGS substantially as the photo-z's obtained with *ugriz* are very good to begin with.

2. For LRGs ($i < 20.5$, $z_{\rm spec} < 0.8$), galaxy size is found to improve photo-z's only marginally, though we see substantial improvement in outliers.

3. For galaxies in the Stripe-82 Sample ($i < 24$, $z_{\rm spec} < 1.5$), although the improvement individual morphological parameters brings to photo-z's is not as significant as the previous two samples, multiple morphological parameters still bring substantial improvement. Most improvement is found for bright, low-redshift objects, and objects which are very small or very large. Galaxy size is found to still improve photo-z's with the presence of infrared bands, and with the help of quasar redshifts in the training, galaxy size would help to significantly reduce the number of catastrophic outliers in a sample with a weak star-galaxy separation process.

4. When tested on a sample reweighted with respect to the CS82 Survey ($i < 24$, $z_{\rm spec} < 1.5$), no significant improvement in photo-z is found when multiple morphological parameters are used in training with 5 *ugriz* bands. However, larger improvement is found under suboptimal conditions, namely when the number of filters in the survey is limited (less than 5 bands), and when the quality of photometry is low.

Ultimately, we find that it is safe to include morphological parameters like surface brightness, radius, axis-ratio and the Sérsic index for machine learning photo-z estimation, regardless of the depth of the sample. We look forward to see applications of this work in upcoming surveys like LSST and Euclid, which would provide us with photometry and morphology of higher quality.

7

PROJECT 2: PHOTOMETRIC REDSHIFT PDFS AND DISTRIBUTIONS

Remember your Creator in the days of your youth, before the days of trouble come and the years approach when you will say, "I find no pleasure in them" – before the sun and the light and the moon and the stars grow dark, and the clouds return after the rain...

Ecclesiastes 12:1-2

7.1 INTRODUCTION

In Chapter 6 we have studied the impact of morphology on photo-z point estimates. However in recent years the use of photo-z probability distributions (PDFs, also known as $p(z)$) have become common, as it is found that they hold more information to redshift than its mode or mean. Besides, it was also found that the stacked $p(z)$ of an entire galaxy sample produced better redshift distributions $n(z)$ which are widely sought after especially in weak lensing analysis. The reader could refer to Section 3.3.2 where we have discussed the development and improvement of $p(z)$ and $n(z)$ in more detail.

In this chapter we turn our focus from photo-z point estimates to PDFs and distributions, where we conduct analyses to improve and better characterise $p(z)$ and $n(z)$. This chapter is divided into 3 main sections. In Section 7.2 we continue our motivation from the previous chapter, we study the effects of galaxy morphology on the individual $p(z)$ of galaxies and the overall $n(z)$ of the CS82 Sample by using ANNZ2. In Section 7.3 we apply the results from the previous section and Section 6.5 to produce the photo-z catalogue for the CS82 Survey. Finally in Section 7.4, we compare the performances of ANNZ2 and DELIGHT, two state-of-the-art photo-z algorithms in point estimates, $p(z)$ and $n(z)$ on the Buzzard Simulated Galaxy Catalogue in the context of LSST. As the efforts to quantify $p(z)$ and $n(z)$ are relatively new in among the photo-z community, this chapter also serves as an exploratory work into finding the suitable metrics to quantify these attributes.

Here we note that the work in Section 7.3 was published in Soo et al. (2018). Parts of the work in Section 7.2 are also published in the same paper, however the analysis and content published has been heavily extended with the use of new metrics, therefore we see the contents in this section as an extension to the work of Soo et al. (2018). Finally, Section 7.4 is part of a collaborative work with the LSST-DESC PZ Working Group, where a paper which quantifies the $p(z)$ and $n(z)$ of various photo-z codes on the Buzzard Sample (in the context of LSST) is to be published late 2018 (Schmidt et al., in prep.). For the purpose of that paper, we have

submitted our photo-z results (ANNz2 and DELIGHT) to be compared with the performance of other codes, however we have also obtained permission to utilise the photo-z's produced for personal analysis, which result in the contents of Section 7.4. While we may mention briefly and qualitatively the motivation and background of the work of Schmidt et al. (in prep.), we assure the reader that the analyses presented in Section 7.4 are original and not the work of others.

7.2 IMPACT OF MORPHOLOGY ON P(Z) AND N(Z)

7.2.1 *The CS82 Sample*

In Section 6.5 we have seen how morphology improves photo-z point estimates produced by ANNz2 using various morphological parameters in different conditions. However, the full photo-z posterior distribution provides more information and is frequently used to estimate sample redshift distributions. In this section we are interested to know if galaxy morphology affects the shape of the individual $p(z)$ and the $n(z)$ of the entire sample.

To start, we first assume that an improvement in the photo-z estimates implies an improvement in the $p(z)$ and $n(z)$. In the context of CS82, we have however seen from Section 6.5 that the improvement when using morphology (μ, r, q, p and n) with 5 *ugriz* bands is marginal. Therefore in this section we chose to use the results of Section 6.5.2 as the sample of our study due to the higher improvement percentages observed. This was the sample where we compared the changes in metric when morphology is added to a number of bands less or equal to 5 (e.g. i, ri, grz and etc). In addition to that, we consider both the unweighted and weighted cases to see if there are differences in results between the two. ANNz2 will be the primary algorithm used to produced the $p(z)$ and $n(z)$ throughout this section.

7.2.2 *Performance of Odds Parameter on ANNz2*

As discussed in Section 3.3.2, BPZ introduced a Bayesian odds parameter (which is denoted by Θ in this work) to characterise the 'peaky-ness' of the $p(z)$, and it is represented in Equation 71. The odds parameter holds a value between 0 and 1, it has been used by BPZ as a quality cut to keep the more reliable photo-z's and PDFs (Benítez 2000).

Taking this idea from BPZ, we calculated Θ for every $p(z)$ estimated by ANNz2 for the CS82 Sample, and here we assess the performance of this newly incorporated parameter to see if it would act as a suitable quality cut for our sample. When training with *ugriz*, we find that an 80% completeness cut in Θ (i.e. discarding 20% of objects with the lowest Θ) retains more than 70% of objects with $z_{\rm phot} > 0.9$, while only slightly more than 50% are kept when using the same completeness cut in photo-z error (Figure 50). This shows that with Θ cuts we get to keep more objects from the higher redshift regime than with ANNz2 photo-z error, which is therefore a more suitable quality cut to be used.

We also evaluate the performance of ANNz2 by comparing its photo-z with those produced by BPZ. We used default settings and the CWW template, and we compare both photo-z's in Figure 51. From the figure, we see that ANNz2 is performing better than BPZ in terms of the number of outliers when no cuts are taken. We also applied an 80% completeness cut using

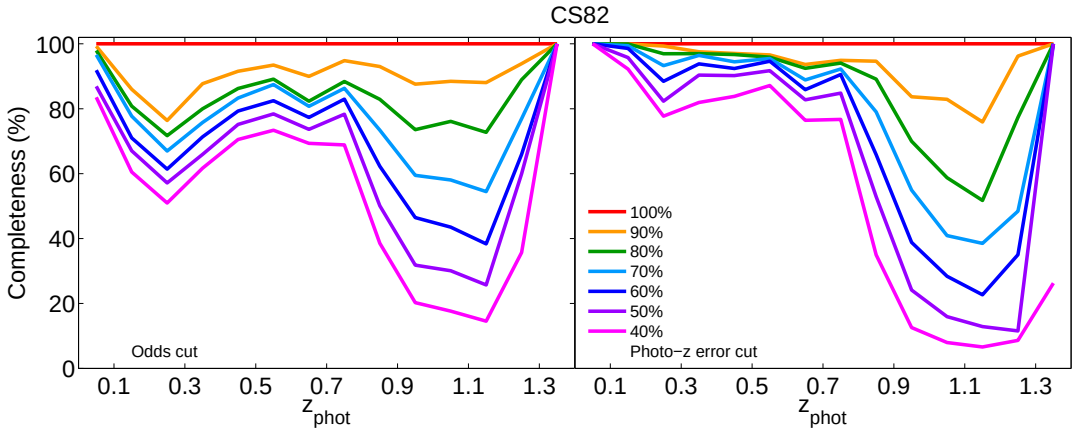

Figure 50: A comparison between the use of the odds parameter (left) and the photo-z error (right) as a quality cut for ANNz2, with respect to z_{phot} at different completeness levels (with respect to the number of objects in the individual z_{phot} bins).

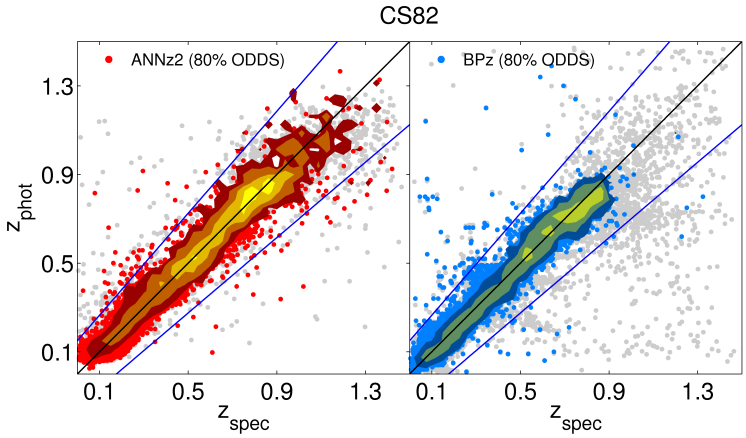

Figure 51: Photo-z v.s. spec-z comparing the performance of ANNz2 and BPz of the CS82 Sample. The coloured dots show objects with an 80% odds cut.

the respective odds values for ANNZ2 and BPZ (shaded in colour), and we find that the photo-z's produced by ANNZ2 not only efficiently reduce the number of outliers, but also keep more objects with higher redshift ($0.9 < z < 1.5$), which BPZ discarded almost completely.

The results of ANNZ2 produced a bias at very low redshift (a visible small gap in the lower left corner of Figure 51). This is a result of the reweighting, as the machine learning algorithm highly down-weighted bright and low-redshift objects, putting more emphasis on the higher redshift objects. The galaxies affected by this bias occupy a very small fraction of the survey volume; thus its effect is negligible across the general metrics. Besides, most objects in this low-redshift region have good spectroscopic redshifts available to be used, and they could also be removed by using a magnitude cut of $i > 20$ for lensing studies.

7.2.3 *Impact on p(z)*

We first evaluate the impact of galaxy morphology on the quality of $p(z)$. To do so, we employ 2 metrics, namely the mean CRPS value ρ_{CRPS} and the root-mean-square PIT value ρ_{PIT}. These metrics are fully described in Section 5.7.2, but as a quick summary, ρ_{CRPS} tells us on average how well the $p(z)$ encapsulates the true redshift (the smaller the value the better), while ρ_{PIT} tells us on average if the sizes of the $p(z)$ width produced are adequate (large value means that on average the $p(z)$ are either too wide or too narrow).

At the same time, we also explore the effects of galaxy morphology on the following 3 values: (1) the mean odds value $\bar{\Theta}$, to see if the reliability of the $p(z)$ increases; (2) the mean height of the $p(z)$ maximum peaks $\bar{h}$, for which we assume the higher the better; and (3) the mean 68th percentile width $\bar{w}_{68}$ of the $p(z)$, where we expect the width to decrease with morphology. We note here however that we do not expect these values or metrics to correlate with one another, especially since a high ρ_{CRPS} (accurately peaked) does not imply a low ρ_{PIT} (correctly shaped).

The values of ρ_{CRPS}, ρ_{PIT}, $\bar{\Theta}$, $\bar{h}$ and $\bar{w}_{68}$ are calculated for both the unweighted and weighted CS82 samples, for cases when trained with different numbers of magnitudes with and without morphology, and the results are shown in Figure 52. Overall for both cases we see that most metrics experience more improvement than degradation when morphology is added to training, and the trend of less improvement in the case of larger number of bands is also seen in most metrics. Notably ρ_{CRPS}, we see that galaxy morphology improves the forecast of the $p(z)$ immensely in both the weighted and unweighted samples, in most cases improvement of at least 5% is achieved.

The change in ρ_{PIT} however is different from that of ρ_{CRPS}, where we see improvement in the unweighted sample, however when the training is weighted the metric degrades. The reason for an increase in ρ_{PIT} here by looking at the metric itself is not easy to pin down, since it could go both ways: either the widths of the $p(z)$ are going too narrow, or too wide. But with the values of $\bar{w}_{68}$ generally decreasing, we deduce that the sizes of the $p(z)$ have turned narrower than they should be with the inclusion of morphology.

To get a clearer picture for the change in ρ_{PIT} would be to see the QQ plots, as shown in Figure 53 for the cases of i, grz and $ugriz$ with morphology. It is evident from the QQ plots that in fact the $p(z)$ widths are well distributed across the entire sample, since both the green and red lines are very close to the diagonal line, seemingly indistinguishable. Therefore it is hard

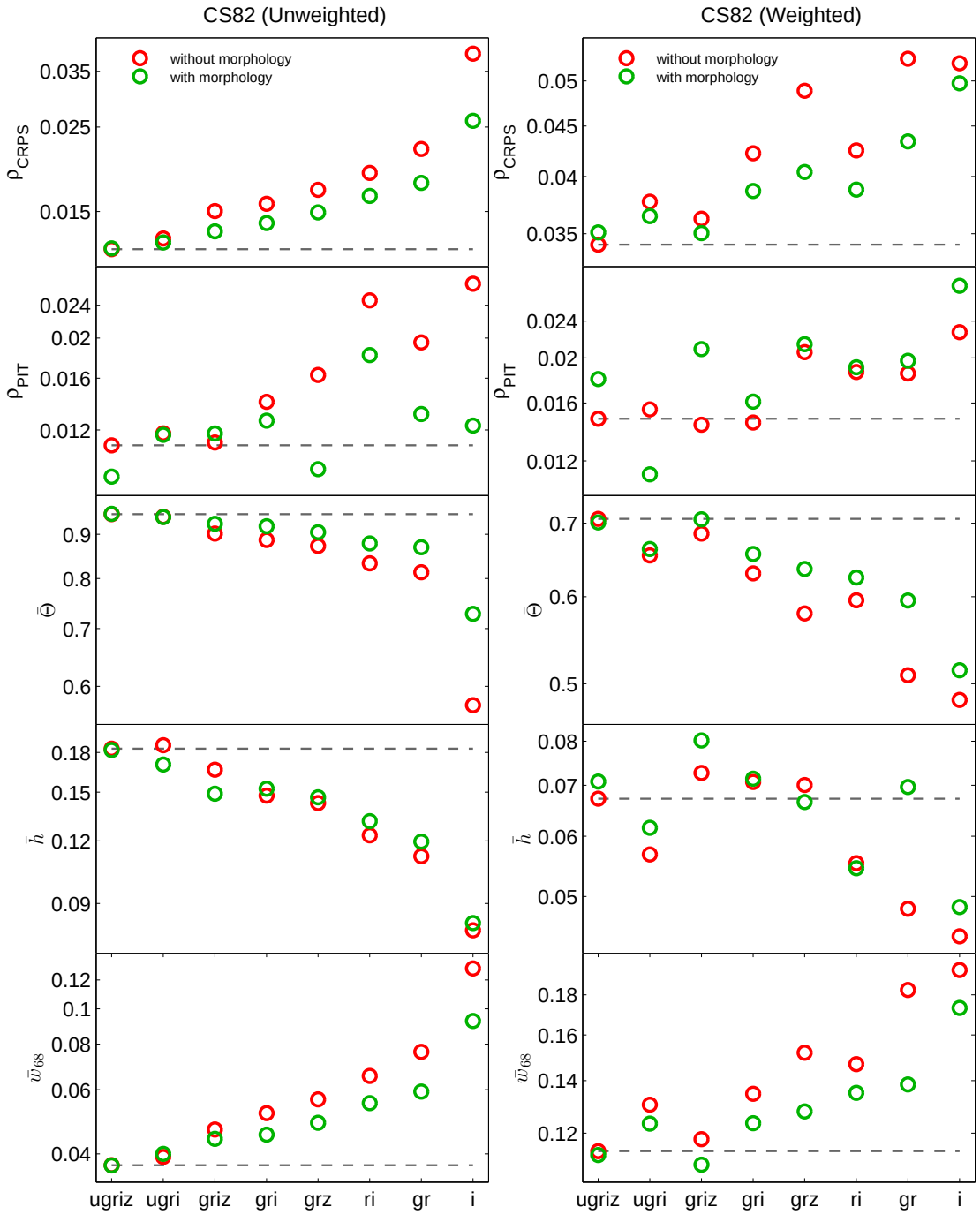

Figure 52: Comparison of the mean CRPS (ρ_{CRPS}, first row) and root-mean-square PIT value (ρ_{PIT}, second row) when morphology is added to the training (green circles) on the unweighted (left) and weighted (right) CS82 training sample. Also shown are the mean odds ($\bar{\Theta}$, third row), mean peak height ($\bar{h}$, fourth row) and mean 68% width ($\bar{w}_{68}$ last row) of the $p(z)$ produced for each case. The grey lines indicate the metric value of the pure *ugriz* run. The definition of these metrics can be found in Section 5.7.2.

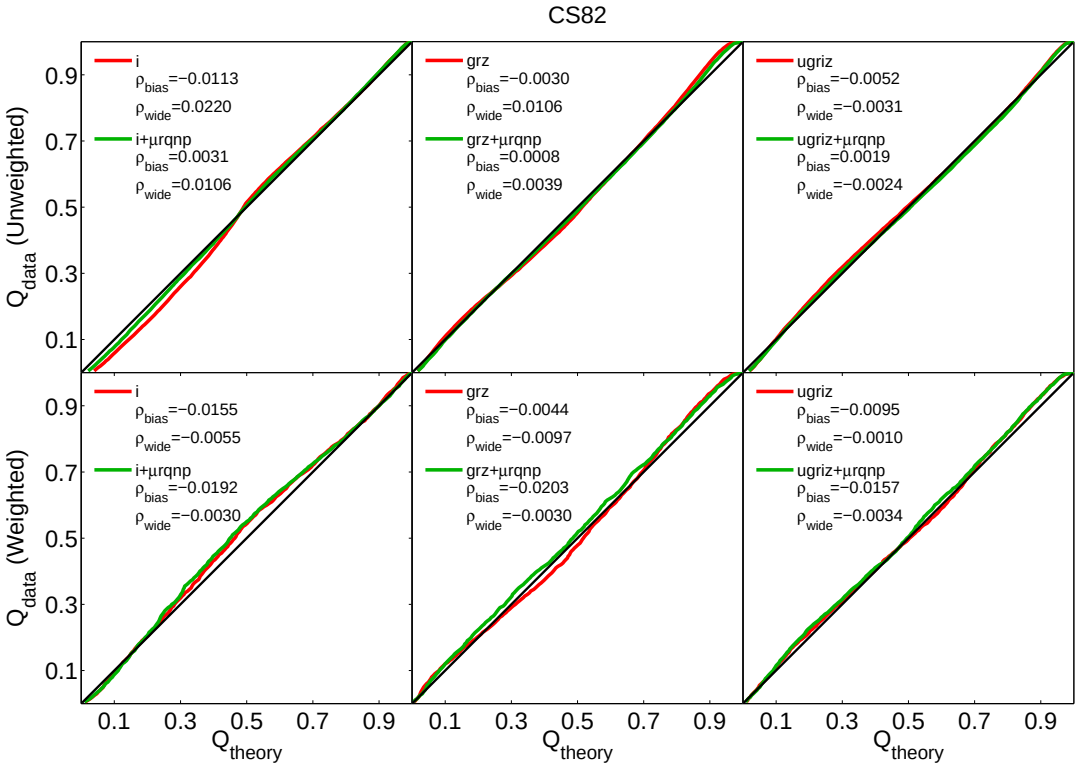

Figure 53: *QQ* plots for the photo-*z*'s of the CS82 Sample when trained with *i* (left), *grz* (middle) and *ugriz* (right), both with and without morphology (green and red) for the unweighted (top) and weighted (bottom) samples respectively.

to identify any traits of overprediction / underprediction or the $p(z)$'s being too narrow or too broad, for both the weighted or unweighted cases just from the QQ plot alone. However with the help of ρ_{bias} and ρ_{wide}, we can identify the slight over- / under-predictions and overall wideness of the PDFs. As with ρ_{PIT}, we see a slight improvement in ρ_{bias} and ρ_{wide} for the unweighted sample when morphology is added, but not for the case when the metrics are reweighted with respect to the CS82 Target Sample. Therefore it is clear that ρ_{PIT} is not a good metric for the $p(z)$ in the context of CS82 since any improvement / degradation shown in the tables are insignificant in reality. It is difficult to conclude the effects of morphology on the CS82 QQ plots, since the overall quality of the PDFs as shown are very high to begin with.

In regards to $\bar{\Theta}$, for a 5 *ugriz* band training with morphology we find that the change in $\bar{\Theta}$ is almost negligible with morphology in the unweighted case: $\bar{\Theta}$ increased from 0.950 to 0.951 when multiple morphological parameters are included, partly because $\bar{\Theta}$ is very high to begin with. However we see a general improvement in both the unweighted and weighted case when morphology is added to the training, which indicates that morphology is indeed helping the ANN to improve the confidence in photo-z values. We also see a direct correlation with ρ_{CRPS} as expected, since both measure how well the expected redshift has been encapsulated within the PDF. However when checked for individual galaxies, we have verified that there is almost no correlation between the improvement in photo-z and the improvement in Θ for individual galaxies when morphology is added, even for the case of *i*-band. Therefore this suggests that although a high Θ does not necessarily dictate the quality of an individual galaxy's photo-z, it is sufficiently useful to remove outliers across an entire sample of galaxies.

The change in $\bar{w}_{68}$ follows a similar trend as $\bar{\Theta}$, where we see that morphology on average has successfully reduced the widths of the $p(z)$. However, the impact of morphology on $\bar{h}$ shows a different story: there is a mixture of improvement and degradation. In fact, we find low correlation not only between $\bar{h}$ and the performance metrics across number of filters used, we also find low correlation between the improvement with $\bar{h}$ and the improvement in photo-z for individual objects when morphology is added. We also do not see significant correlation between the improvement of h and Θ either.

To understand this further, we take the same case studies on *i*, *grz* and *ugriz* and tabulate the distributions of Θ, w_{68} and h for both the unweighted and weighted samples as shown in Figure 54. From both the unweighted and weighted case, we see that the change in height of the $p(z)$ is in fact very small, and this is in contrast with the change in width and odds. This signifies that the height of the $p(z)$ is not a good indicator for improvement in $p(z)$, and suspect that the improvement / degradation may be due to random noise in the training. From the figure we also see visible shifts in distributions of w_{68} and Θ (especially for the unweighted case of *i*), confirming our deduction that morphology in general improves these parameters overall. The exception of course is the case of *ugriz*, where we have established in the previous chapter that morphology has insignificant impact on it.

The motivation to study the change in $\bar{h}$ was to see if morphology could decrease the degeneracy of multiple peaks, making only one peak stand out to obtain a more accurate photo-z. However we deduce that the impact morphology brought is not high enough to do so. This can be further understood by looking at a few sample $p(z)$ of both samples. Figure 55 and Figure 56 show the sample $p(z)$ for the unweighted CS82 training sample and the sample weighted with respect to the CS82 target set, respectively. As shown in both figures, the $p(z)$ produced

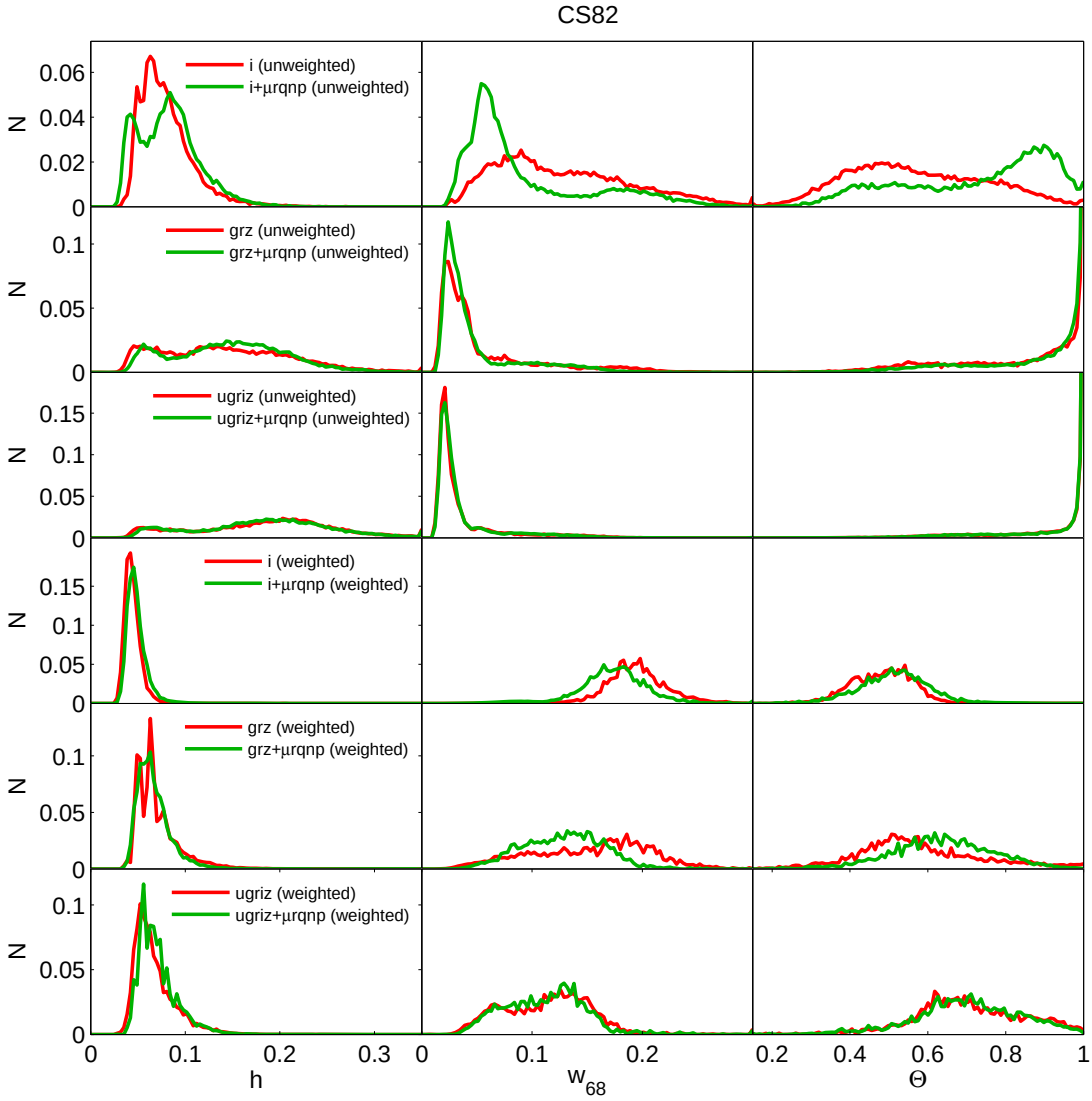

Figure 54: The distribution of the $p(z)$ height h (left), 68% width w_{68} (middle) and odds parameter Θ when trained using *i*, *grz* and *ugriz*, with and without morphology (green and red lines). The top three panels show the results for the unweighted CS82 training sample, while the bottom three when reweighted to the target set.

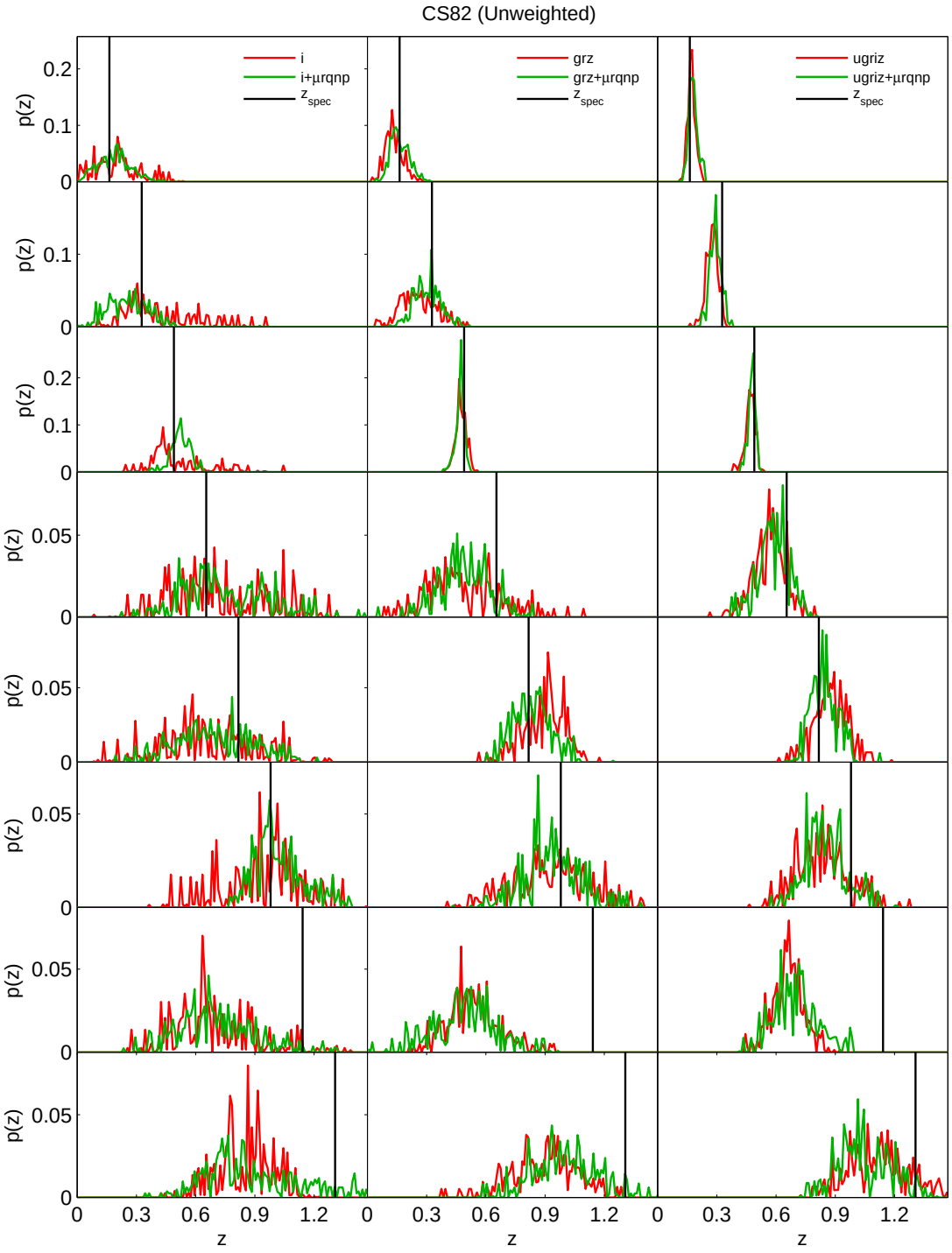

Figure 55: Sample $p(z)$ for the CS82 Sample when trained with *i* (left), *grz* (middle) and *ugriz* (right), with and without morphology (green and red), produced by ANNz2.

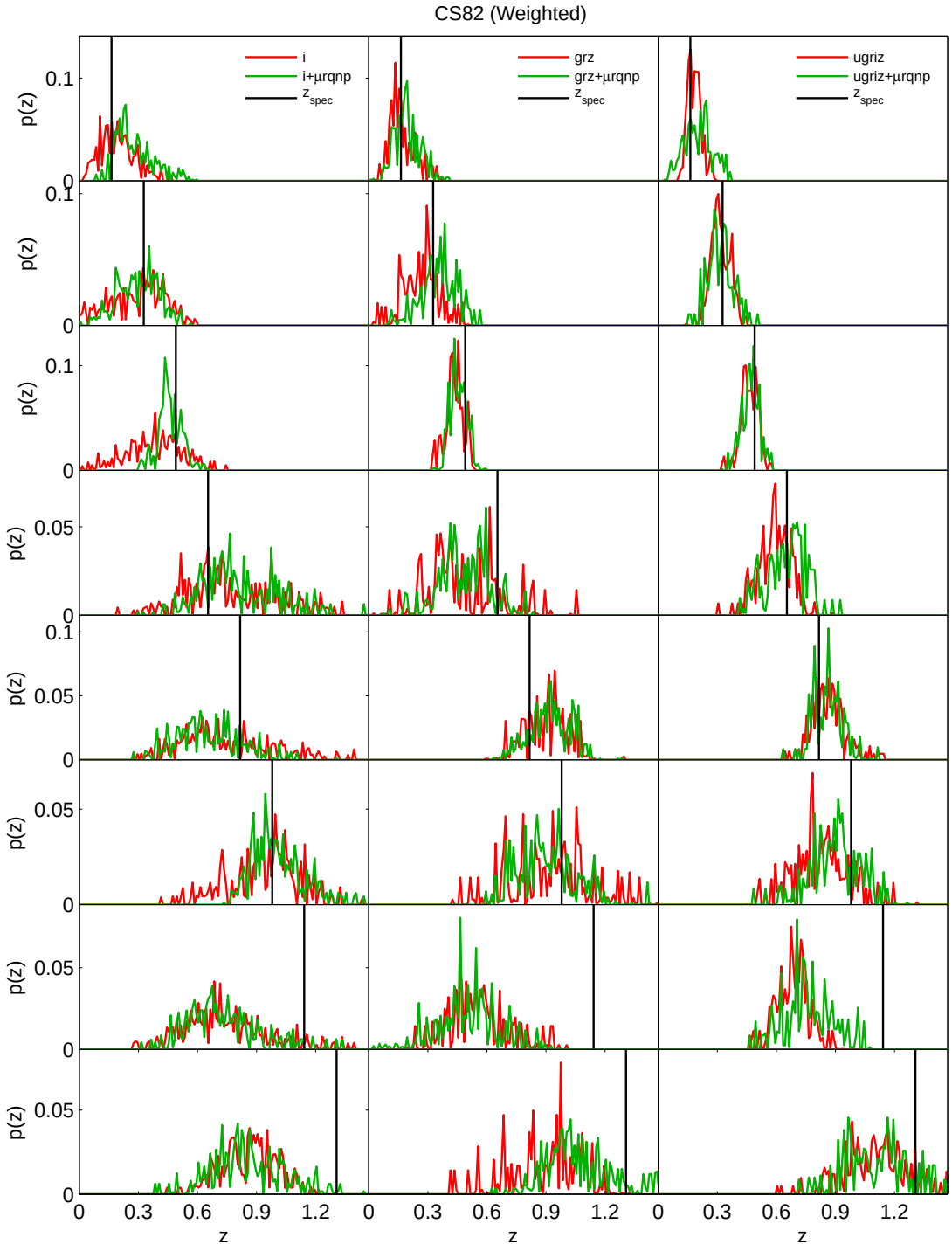

Figure 56: Sample $p(z)$ for the CS82 Sample when trained with reweighting with respect to the CS82 Target Set, produced by ANNZ2 when trained with *i* (left), *grz* (middle) and *ugriz* (right), with and without morphology (green and red). Note that these are the same objects as those in Figure 55.

by ANNZ2 in fact have very fuzzy shapes, unlike most template-based codes, due to the way the machine learning algorithm was tuned to work. Not having a visible peak and the presence of multiple small peaks renders the study of h not meaningful, at least for the context of ANNZ2. The $p(z)$ also explains the reason why the mean photo-z of the distribution is used instead of the peak as the best point estimate for ANNZ2. However as we have only used 5 committees for each ANNZ2 run in this work (see Section 5.3.4 for the hyperparameter settings of ANNZ2 used in this work), we note that the PDFs would appear smoother if a higher number of committees are used, though we believe that the impact on the position of the peaks would be marginal. Both figures also illustrate the reweighting scheme that we used, showing that the $p(z)$ in general have their peaks slightly closer to their respective true redshifts (black line) at $z_{\rm spec} > 0.7$ since objects in this regime have higher weights and should be more accurate than their unweighted counterparts.

7.2.4 *Impact on n(z)*

In this section, we are interested if galaxy morphology improves the performance of the $n(z)$ of the sample. Using the same two samples training with and without morphology, we obtain the $n(z)$ for each case by stacking the individual $p(z)$, and compare this with the distribution of $z_{\rm spec}$. We use two metrics to evaluate the performance of the $n(z)$: the KS statistic $n_{\rm KS}$ and the root-mean-square of the difference between the spec-z distribution $n_{\rm spec}(z)$ and the photo-z distribution $n_{\rm phot}(z)$, which we denote as $n_{\rm RMS}$. The full definitions of these metrics are described in Section 5.7.3. We calculate the values of both of these metrics for the unweighted and weighted sample. We calculate the percentage differences for these metrics when morphology is included, and the results are tabulated in Table 22.

From the table, it is evident that we do not see any trends in improvement and degradation in either $n_{\rm KS}$ or $n_{\rm RMS}$ when morphology is included or when the number of filters varied. While the change in $n_{\rm KS}$ are mostly negative, the change in $n_{\rm RMS}$ are relatively small, and mostly at percent level for the weighted sample. For a testing size of 19 833, the KS test null hypothesis is rejected at 5% level when $n_{\rm KS} > 0.01366$. From the table, we see that any training in the unweighted sample with more than 2 bands have passed the test, however none of the trainings in the weighted sample passed.

The failure of the KS test on the weighted CS82 Sample can be understood by looking at the $n(z)$ distributions in Figure 57, where we show the distribution of the spectroscopic redshift (pink histogram), and compare the stacked $n(z)$ for the photo-z runs of i, grz and $ugriz$. As discussed in Section 5.7.3, the KS statistic tells if two distributions are drawn from the same distribution, and it is sensitive to the maximum change in the distribution. As seen from the figure, the weighted spectroscopic redshift distribution, unlike the unweighted one, is very uneven and features many spikes even with double the bin size, this is caused by the weights acting on the training objects, which in fact is a relatively small sample. We expect the reweighted $z_{\rm spec}$ distribution to be smoother if the training set was larger, which would result in lower $n_{\rm KS}$ values. Therefore we do not recommend the use of the KS test as a metric to characterise redshift distributions, especially in weighted training samples, while $n_{\rm RMS}$ is a more suitable metric to be used.

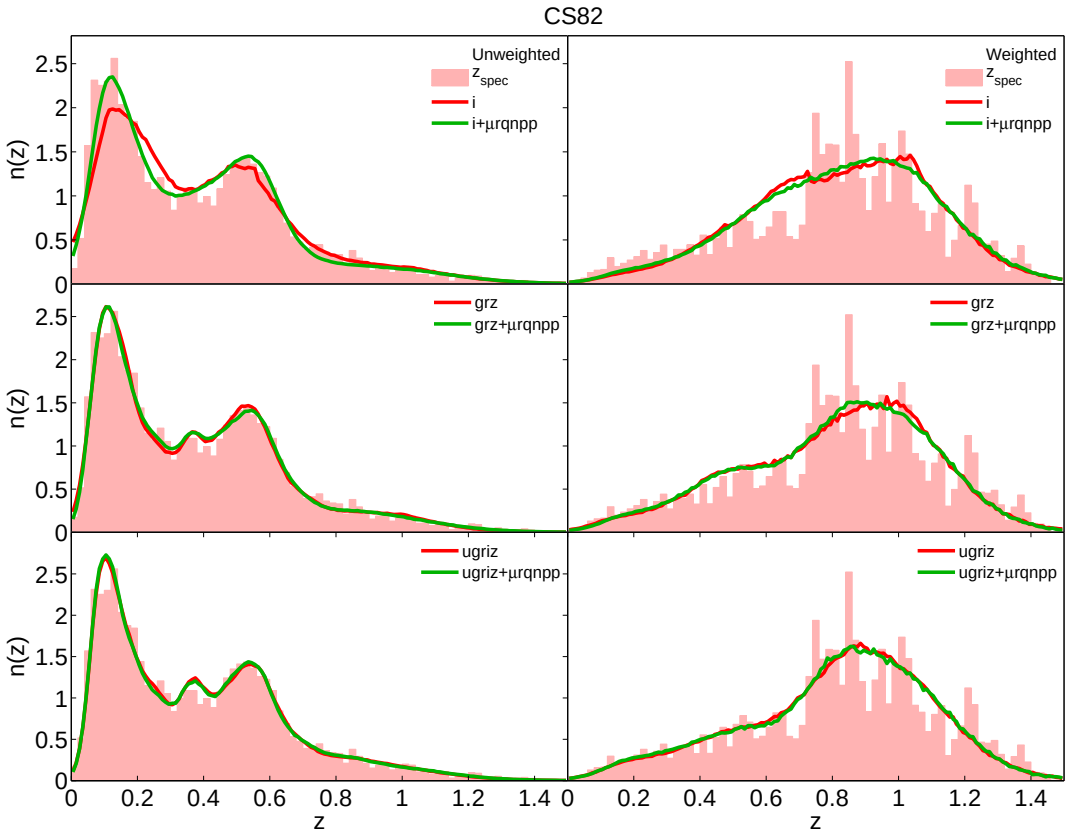

Figure 57: The comparison between the $n(z)$ distributions of the spec-z (pink histogram) with the photo-z's trained with i (top), grz (middle) and $ugriz$ (bottom), respectively, with and without morphology (green and red). The left and right panels show the unweighted and weighted samples respectively.

Table 22: Relative improvement in KS statistic (n_{KS}) and root-mean-square value of $n_{\mathrm{spec}}(z) - n_{\mathrm{phot}}(z)$ (n_{RMS}) for both the weighted and unweighted CS82 Sample, trained with and without morphology for different number of bands. The definition of these metrics can be found in Section 5.7.3.

Inputs	**Unweighted Sample**				**Weighted Sample**			
	n_{KS}	$\Delta\%$	n_{RMS}	$\Delta\%$	n_{KS}	$\Delta\%$	n_{RMS}	$\Delta\%$
i	0.0519		0.222		0.0450		0.362	
i+morph.	0.0221	57.4	0.154	30.4	0.0374	16.9	0.347	4.1
gr	0.0166		0.145		0.0529		0.353	
gr+morph.	0.0156	6.4	0.150	−2.9	0.0391	26.2	0.333	5.7
ri	0.0125		0.141		0.0417		0.327	
ri+morph.	0.0141	−12.4	0.135	4.2	0.0438	−5.1	0.326	0.5
gri	0.0094		0.137		0.0372		0.325	
gri+morph.	0.0134	−42.2	0.135	1.8	0.0407	−9.4	0.323	0.5
grz	0.0113		0.140		0.0410		0.340	
grz+morph.	0.0120	−6.3	0.136	3.4	0.0412	−0.6	0.330	3.1
griz	0.0141		0.147		0.0375		0.324	
griz+morph.	0.0157	−10.9	0.144	2.0	0.0419	−11.7	0.325	−0.2
ugri	0.0116		0.131		0.0406		0.322	
ugri+morph.	0.0108	6.4	0.140	−6.9	0.0394	3.1	0.322	0.0
ugriz	0.0121		0.141		0.0423		0.321	
ugriz+morph.	0.0113	6.8	0.141	0.0	0.0410	3.1	0.316	1.6

The greatest takeaway from Figure 57 however is that the $n(z)$ produced for *ugriz* and *grz* are very similar, only distinguishable by small features for both the weighted and unweighted cases. Upon checking the values of n_{RMS} from Table 22, we find that in fact the $n(z)$ produced using 2 filters with morphology only differs from the $n(z)$ produced using 5 filters by no more than 5% in value for n_{RMS}. The $n(z)$ for $i + \mu rqnp$ performs arguably well for $z_{\mathrm{phot}} < 0.5$ and $z_{\mathrm{phot}} < 1.0$ in the weighted sample. This result is mainly due to the redshift-magnitude relation in the reweighting and the optimisation to produce better $n(z)$ by ANNZ2, nevertheless it is still good news for surveys with limited number of filters, allowing them to produce good $n(z)$ with less than 5 filters.

From this section we conclude that galaxy morphology improves the $p(z)$ through the ρ_{CRPS}, and it also improves the mean odds values and reduces the widths of the PDFs in cases with less than 5 filters. As for $n(z)$, galaxy morphology does not improve the photo-z redshift distributions, however ANNZ2 is capable of producing high quality $n(z)$ by merely using 2 bands and several morphological parameters. The effects of morphology on $n(z)$ will be further discussed and visualised in the next section as the results for the CS82 Photo-z Catalogue unfold.

7.3 THE CS82 PHOTOMETRIC AND MORPHOLOGICAL REDSHIFT CATALOGUE

We now utilise the results we have obtained from the previous section and chapter to produce a photo-z catalogue for the CS82 Survey. As mentioned in Section 4.1.4, the CS82 Morphology Survey only has one *i*-band magnitude, and we seek to use multiple morphological parameters,

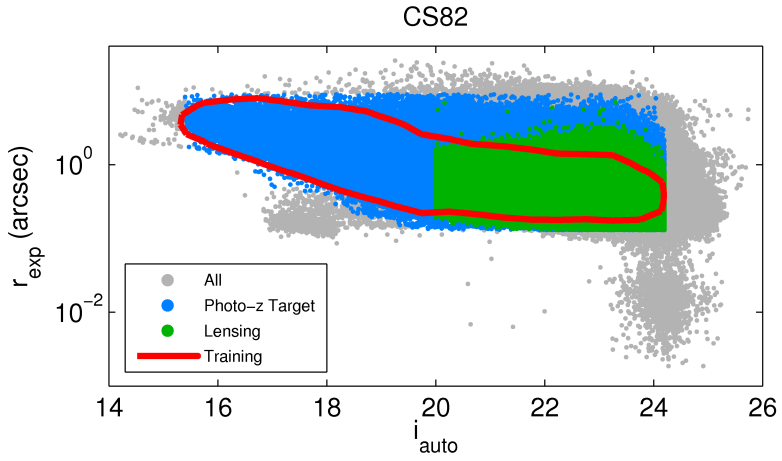

Figure 58: Plot of exponential radius (r_{exp}) v.s. Kron i-band magnitude (i_{auto}) for the CS82 Sample, comparing our training sample (red contour) to the CS82 Target Sample (blue), its lensing subsample (green), and the full CS82 Morphology Catalogue (grey).

SDSS *ugriz* photometry when available, and otherwise the CS82 i-band magnitude, together with quasar redshifts to produce photo-z's for galaxies.

To produce this photo-z catalogue, we use a spectroscopic training set similar to the one used in Section 6.5, however this time with the addition of quasar spectra (motivated by the results from Section 6.4.2, where the inclusion of quasar spectra in training improves the photo-z quality of the catalogue). Representability is solved by the same reweighting scheme as described in Section 6.5, but we also ensure that the target sample has the same photometric and morphological cuts and limits as the training sample. The photometric sample selection is the same as the CS82 Sample, this results in a training set of 64 591 objects and a target set of 5 777 379 objects in which photo-z's will be estimated.

For weak lensing analyses, a subset of the target set has been selected for further evaluation, referred to as the *lensing subset*. This subset has extra sample selections as follows: the objects should have good Lensfit shape measurements (`WEIGHT>0`, Miller et al. 2013), and be classified as galaxies[1] (`FITCLASS=0`). This subset has 3 536 783 objects. The size-magnitude diagram of the training and target sample is shown in Figure 58.

Similar to the Section 6.5, the photometry in our training set is cross-matched to spectroscopic redshifts from SDSS, BOSS, DEEP2, WiggleZ and VVDS. The distribution of the training and target sets in colour-magnitude space are illustrated in Figure 59. We further show the breakdown of the objects in the training set (coloured in red) with respect to the sources of their spectroscopic redshifts and their relative densities, where we see that the different surveys contribute to different regimes in the colour-magnitude diagram. The figure also highlights the importance of reweighting, since the target set is obviously weighted differently from the combined training set used.

The training sample is used to produce photo-z point and PDF estimates for the CS82 Morphology Catalogue by training SDSS *ugriz* with CS82 morphology $\mu rqnp$ to be tested on objects

1 In particular, this limits the sample to $i > 20$ as brighter galaxies tend to be too large to be processed efficiently by the Lensfit algorithm.

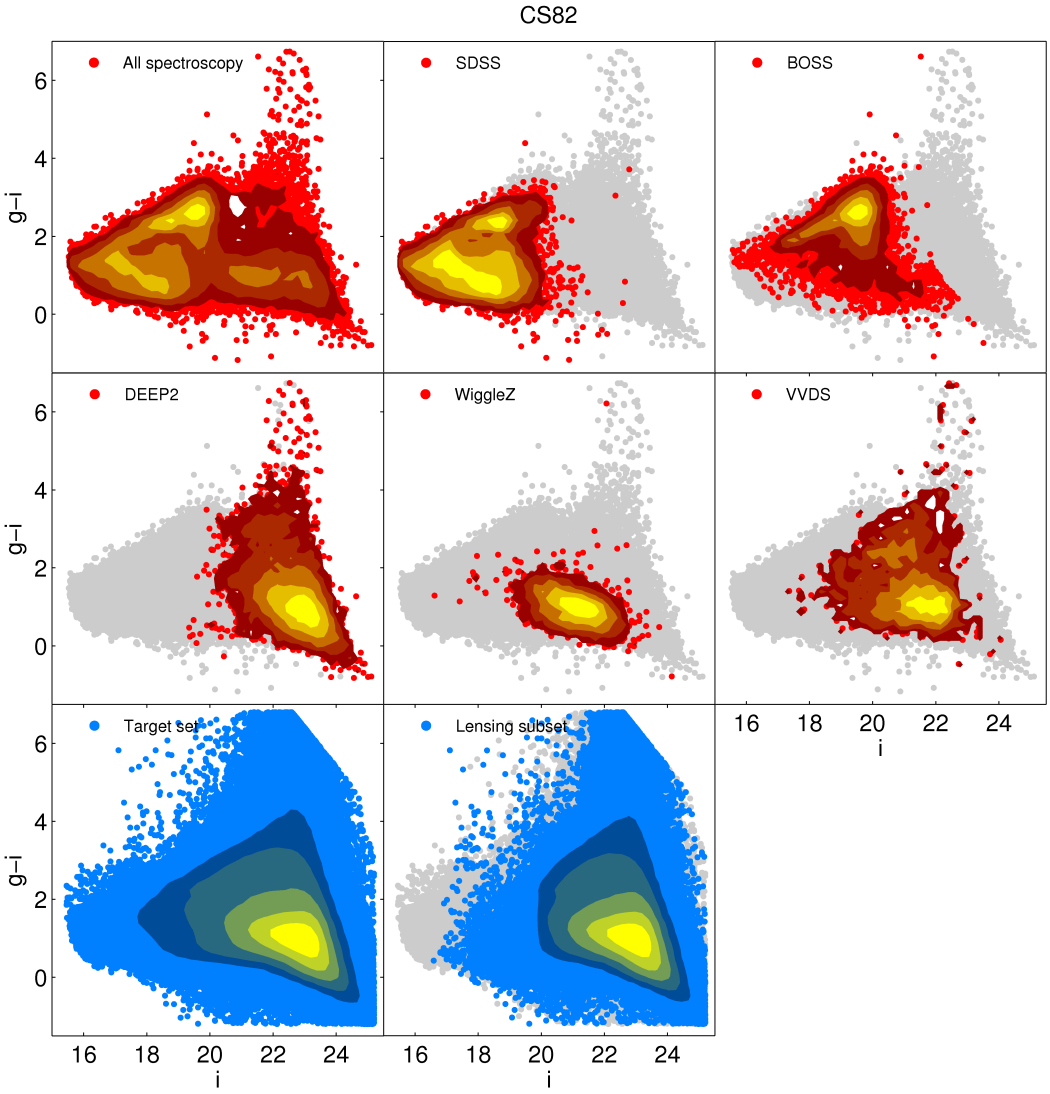

Figure 59: Apparent *g*-*i* vs *i* magnitude for objects from the CS82 training sample (red), and part of the CS82 Target Sample with matched SDSS photometry (blue) used in this study. The 6 panels on the top highlight the training objects according to their spectroscopic sources (SDSS, BOSS, DEEP2, WiggleZ and VVDS), while the bottom right plot highlights the lensing subsample.

Table 23: Comparison of performance between photo-z's produced by Reis et al. (2012), BPZ, EAZY and ANNZ2 (our results) for the testing set of the CS82 Photo-z Catalogue. Note that these metrics have been weighted in accordance to the densities of the target sample, and no odds cuts have been applied.

Photo-z Code	$\sigma_{\rm RMS}$	σ_{68}	$\eta_{\rm out}(\%)$
Reis et al. (2012) (ANN)	0.0966	0.0677	8.26
BPZ	0.1473	0.0911	19.46
EAZY	0.1234	0.0867	11.89
ANNZ2 (*ugriz*)	0.0915	0.0606	5.97
ANNZ2 (*ugriz*+morph.)	0.0872	0.0583	5.15
ANNZ2 (*i*+morph.)	0.1366	0.1065	15.89

that have cross-matched SDSS photometry in the target set. However, approximately 18% of objects in the target set do not have SDSS *ugriz* photometry, and these objects make up the majority of the sample at $i > 23.8$. Therefore we estimate a second redshift value called *morphological redshifts* ('morpho-z'), allowing the remaining 1 million objects in the catalogue to have redshift estimates despite not having colour information.

The morpho-z is be obtained by training 7 parameters: the CS82 Kron i-band magnitude, the PSF i-band magnitude, and the morphological parameters $\mu rqnp$. We have included the CS82 PSF i-band magnitude in training, as we think that when trained together with the Kron i-band magnitude it would act as a proxy for the spread model of the object. However we note that the difference in results is negligible even when it is not included in training.

Due to the lack of colour for the morpho-z case, we do the reweighting with respect to the i-band and radius r instead, and we find it performs comparably to reweighting in i and g-i, obtaining almost the exact same results as Figure 42. In situations where we have multiple redshift values for each object, we provide a `ZBEST` column in the catalogue – the 'best' redshift for that object, which would be the value of either the spectroscopic redshift, photometric redshift or morphological redshift, in that order.

Point estimate and PDF of photo-z's and morpho-z's are calculated, and the redshift distribution $n(z)$ is produced as well. In order to evaluate the performance of our results, we select a number of objects from the spectroscopic sample as our testing set, and we calculate the weighted performance metrics for each run without any odds cut, which are summarised in Table 23. Our standard *ugriz* photo-z results outperform those by Reis et al. (2012), BPZ and EAZY which we put primarily down to the reweighting. Adding morphology to *ugriz* photometry further improves all metrics for this sample, while the morpho-z's perform similarly to BPZ at least in these global metrics. Note that in order to compare the performance between our photo-z and morpho-z results, the testing set comprises objects with colour information.

Figure 60 shows the photometric v.s. spectroscopic redshift scatter plots for BPZ and our ANNZ2 runs, with the contours being the weighted densities with respect to the target sample. Firstly, we see that with the correct reweighting, ANNZ2 successfully prioritises to obtain better photo-z's for objects between redshift $0.7 < z_{\rm spec} < 1.2$, which we have seen should be the expected peak redshift for the CS82 Photo-z Catalogue in Figure 42. Secondly, we also see that with an 80% odds cut, the training with *ugriz* with morphology outperforms the training with *ugriz* by being

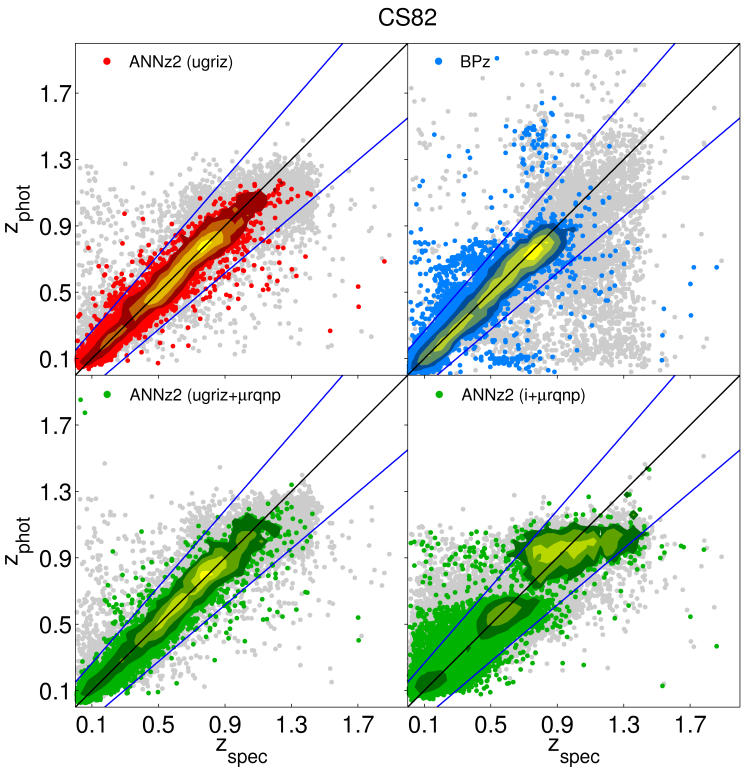

Figure 60: Comparison of photo-z v.s. spec-z for the CS82 photo-z testing set, from top left to bottom right: ANNz2 $ugriz$, BPz, ANNz2 $ugriz + \mu rqnp$ and $i + \mu rqnp$. The grey points show all the objects in the testing set, while the coloured points show the weighted 80% odds cut. Notably the runs trained with morphology keep more of the higher redshift objects than runs trained without morphology.

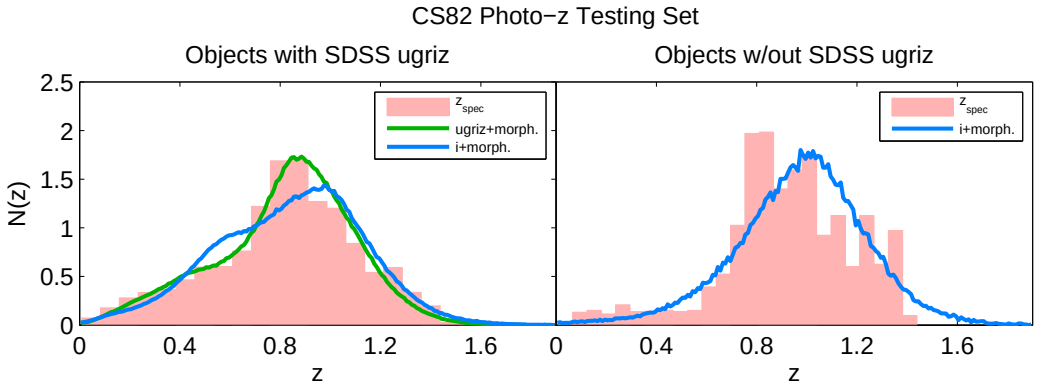

Figure 61: The $n(z)$ distribution of the CS82 photo-z testing set for objects with SDSS *ugriz* magnitudes (left) and those without (right). The reweighted distributions shown are the spectroscopic redshift (pink histogram), *ugriz* with morphology (green) and i with morphology (blue).

able to keep more objects with higher photo-z's, and this strengthens our claim that morphology is a beneficial addition to input for photo-z estimation.

Finally, we evaluate the performance of $n(z)$ for each run in the testing set. In the first panel of Figure 61, we find that the weighted stacked $n(z)$ produced by the $ugriz + \mu rqnp$ run for objects with SDSS *ugriz* magnitudes resembles closely the weighted spectroscopic redshift distribution. The second panel shows an $n(z)$ produced for objects which do not have *ugriz* magnitudes (i.e., no colour information) by stacking the $p(z)$ derived from the i+morphology run. The stacked $n(z)$ is a fair representation of the redshift distribution of objects without colour information, reproducing accurately the number of objects at $z \lesssim 0.6$, and recovering the bulk of the distribution at $z \sim 1$.

We produced a lensing subset of the testing set by applying cuts discussed earlier (resulting in a magnitude range of $20 < i < 24.2$). We find very similar performance for the lensing subsample; see Figure 62 and Figure 63. Note that the upper right panel of Figure 62 corresponds exactly to Fig. 1 of Leauthaud et al. (2017) for direct comparison. The morpho-z's recover the redshift distribution fairly well below $z \sim 0.5$ as well as above $z \sim 1$. These conclusions hold when applying a photo-z cut, e.g. at $z > 0.7$ or $z > 0.8$, which suggests they can be used to define a high-redshift source galaxy bin for galaxy-galaxy lensing applications, with little overlap to be expected for lens galaxy samples at $z < 0.5$.

The photo-z, morpho-z and their respective PDFs produced for the CS82 Survey are publicly available at `ftp://ftp.star.ucl.ac.uk/johnsyh/cs82/`, and the catalogue will be incorporated into the official CS82 website in the future. The individual PDFs for each object for the $ugriz + \mu rqnp$ and $i + \mu rqnp$ are be distributed in separate files from the rest of the parameters. For general purposes, we recommend to use `ZBEST` as the best photometric redshift from this catalogue, an odds cut of $\Theta > 0.5$ is also recommended to remove outliers, which yields $\eta_{\rm out} < 5\%$ and retains 98% of the sample. When using purely the morpho-z values, an odds cut of $\Theta > 0.5$ yields $\eta_{\rm out} < 11\%$ and retains 33% of the sample. To estimate the $n(z)$ of the sample, we recommend the user to stack the individual PDF of `ZPHOT` instead of `ZMORPH`. The description of the headers of table columns are listed in Table 27 at the end of this chapter.

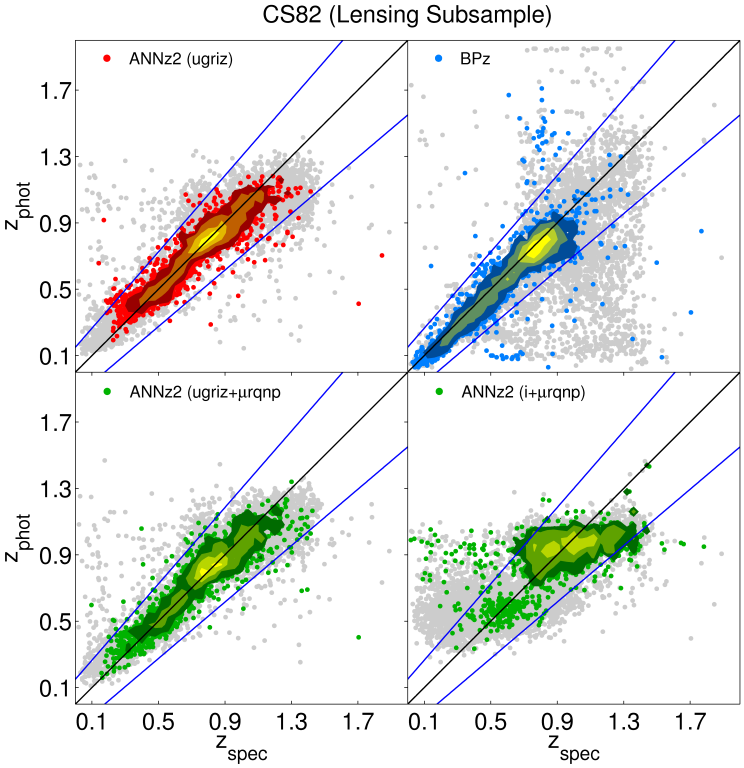

Figure 62: Same as Figure 60, but for the lensing subsample. The top right panel on the left corresponds exactly to Figure 1 of Leauthaud et al. (2017) for direct comparison.

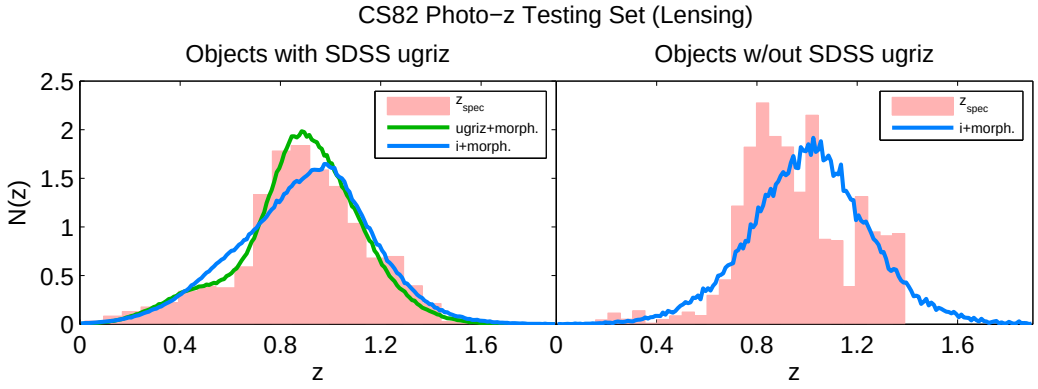

Figure 63: Same as Figure 61, but for the lensing subsample.

7.4 QUANTIFYING P(Z) AND N(Z) IN THE CONTEXT OF LSST

The Large Synoptic Survey Telescope (LSST, Ivezić et al. 2008) is an upcoming large sky survey, it is a wide-field 8.5 m telescope currently being built at Cerro Pachón in northern Chile. The aim of LSST is four-fold: to probe dark energy and dark matter, take an inventory of the Solar System, explore the transient optical sky and map the Milky Way (Ivezić et al. 2008). When the telescope goes online in 2022, it will observe high quality images with superb photometric and astrometric accuracy for 18 000 deg^2 of the sky in *ugrizy* bands (320-1050 nm) for a 10-year operation period, aiming to yield a coadded depth of $r \sim 27.5$.

Since LSST is going to be a photometric survey, photo-z's will play a great role in all dark energy probes to provide the third dimension of the sky map, and therefore require careful planning and calibration. The LSST Dark Energy Science Collaboration Photo-z Working Group (LSST-DESC-PZ) currently works on various projects as part of the LSST Science Roadmap (Robertson et al. 2017) in preparation for photo-z estimation when LSST data is available. Current ongoing projects include providing a testbed for exploring systematics and incompleteness impact on photo-z's (`PZGalaxyGenerator`), optimising photo-z PDF results from various algorithms (`PZPDF`), conducting cross-correlation calibration tests (`PZCalibrate`), and finally, selecting spec-z's from other surveys to be used as training sets (`PZSpeczSelector`).

LSST aims to calibrate and produce photo-z's for the redshift range of $0 < z_{\mathrm{spec}} < 4$ and up to magnitude $r < 27.5$ (The LSST Science Book, Abell et al. 2009). However, for most of the science cases (weak lensing and BAO) the limit would be $i < 25.3$ and $0 < z_{\mathrm{spec}} < 3$, and this high $^S/_N$ subset is known as the *Gold Sample*. The following are the photo-z requirements of LSST:

- $\sigma_{\mathrm{RMS}} < 0.05(1+z)$, but with a goal of 0.02;
- $\eta_{3\sigma} < 10\%$ for all objects, where $\eta_{3\sigma}$ is the 3σ outlier rate of $\frac{z_{\mathrm{phot}}-z_{\mathrm{spec}}}{1+z_{\mathrm{spec}}}$; and
- a photo-z bias less than 0.003, or 0.01 for combined analyses of weak lensing and BAO.

Results from this section are part of a collaborative work with the LSST-DESC PZ working group (`PZGalaxyGenerator`), focusing on characterising and quantifying $p(z)$ and $n(z)$ in the context of LSST, where a paper is to be published by the end of 2018 (Schmidt et al., in prep.). In the following section the background and motivation will be introduced, and we will proceed to use the data products of this work to compare the performance of ANNZ2 and DELIGHT in the context of LSST.

7.4.1 *LSST Photometric Redshift Data Challenge 1*

The LSST Photo-z Data Challenge 1 (DC1) is the first of the 3-part activity which the LSST-DESC PZ group is working on. In DC1, the group focuses on building simulation tools to enable more realistic tests of photo-z algorithms, establishing a framework for running multiple photo-z codes, characterising the photo-z PDFs, and explore efficient ways to store them (Malz et al. 2018).

The first of a series of papers for DC1 is currently being written by Schmidt et al. (in prep.), it focuses on testing and evaluating the performance of multiple photo-z codes on simulations of LSST photometry, with particular emphasis on photo-z PDFs instead of point estimates, which

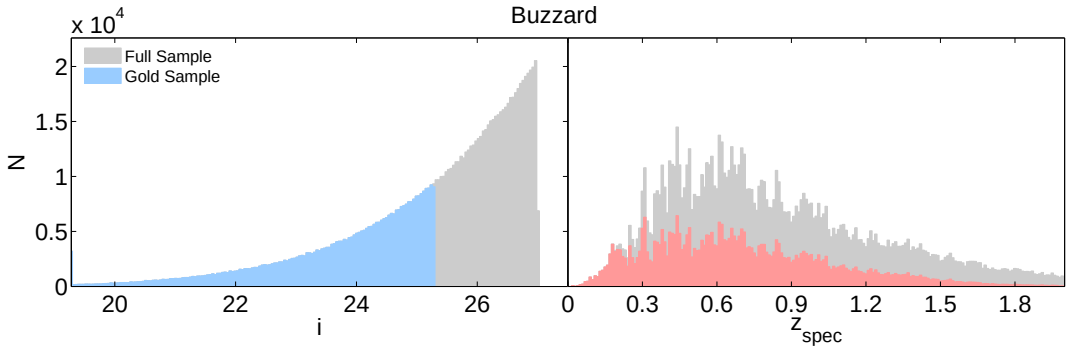

Figure 64: Distributions of the i-band magnitude (left) and $z_{\rm spec}$ (right) for the Buzzard Sample. The Gold Subsample is shaded in colour in the background.

is unprecedented in other current surveys. The Buzzard Simulated Catalogue (DeRose et al, in prep.) underlies the mock catalogues, and a representative training and testing data set was generated for this study (see Section 4.2.1 for full description of the sample). Figure 64 shows the distribution of i-band magnitude and $z_{\rm spec}$ of the Buzzard testing set used. The testing set has a total of 1 000 883 objects with simulated incomplete and missing data (entries removed at random), and the Buzzard Gold Sample ($i < 25.3$) consists of 399 476 objects, which will be the subset on which analyses are conducted.

Multiple photo-z algorithms have participated in this study, the list of codes includes BPZ (Benítez 2000), EAZY (Brammer et al. 2008), LEPHARE (Arnouts et al. 1999), ANNZ2 (Sadeh et al. 2016), the Colour-Matching Nearest-Neighbours photometric redshift estimator (CMNN, Graham et al. 2017), DELIGHT (Leistedt and Hogg 2017), FLEXZBOOST (Izbicki et al. 2017), GPZ (Almosallam et al. 2016b), METAPHOR (Cavuoti et al. 2017b), SKYNET (Graff et al. 2014) and TPZ (Carrasco Kind and Brunner 2013). As the pool of algorithms contains both template based and training based codes, a representative training set of 100 548 objects (of which 43 491 were gold sample objects) and 150 ordered Buzzard templates were given (see Section 5.6.6 for full description of the template set).

As this is one of the pioneering studies on characterising the quality of photo-z PDFs in general, several metrics of performance have been introduced and explored: metrics like conditional density estimation loss, PIT, and QQ plots are used to characterise the individual $p(z)$'s, while the root-mean-square error, KS statistic, Cramer-von Mises statistic, Anderson-Darling statistic and moments were used to characterise the overall $n(z)$ of the sample (see Schmidt et al., in prep. for full details on these metrics). The goal was not to find the best photo-z code; instead the metrics for the photo-z's for each code were compared and analysed to see if they are able to meet the stringent photo-z requirements of LSST. The conclusion of the paper is in the process of being finalized, but overall, the quality of $p(z)$ produced by each code are satisfactory, the point estimates of all codes have reached or are near to the goals of LSST, and future work including analysis of $p(z)$ in tomographic bins, testing on a more realistic sample (incomplete spectroscopic sample, accounting of AGNs), combination of results from different codes and calibrating photo-z's via spatial cross-correlations, have been planned.

We have participated in this work by producing photo-z point estimates and $p(z)$ using ANNZ2 and DELIGHT, where we have made use of the Buzzard templates, training and testing data. The photo-z's and $p(z)$'s have been submitted and compared with other codes. In the following sections, we conduct 3 brief analyses using the photo-z's and PDFs produced by ANNZ2 and DELIGHT. The first two analyses are akin to 'behind-the-scene' analyses on how the photo-z's were produced for the LSST-DESC PZ group: we reveal how we decide on dealing with missing data for ANNZ2 (Section 7.4.2), and the dependence of number of templates used for DELIGHT (Section 7.4.3). Finally in Section 7.4.4 we conduct a full comparison between the performance of photo-z point estimates and PDFs for ANNZ2 and DELIGHT.

7.4.2 *ANNz2 v.s. Missing Data*

One major issue with machine learning photo-z codes is the dealing of missing data. In template-based codes, missing data can be ignored as the fitting is based on interpolation of templates, however it is not the case for machine learning codes: the missing data have to be either replaced by a meaningful value in order for the training and testing to be completed, or an entirely different training and testing session needed to be conducted for objects with similar missing entries.

One usual way to solve the problem is to replace the missing magnitude value with the mean of the column, and increasing the magnitude error by a large factor so that in the training process a lower weight is put on this object. However this method does not work for ANNZ2; since it does not rely on the given magnitude errors in its estimation, the bad entries cannot be downweighted this way. Other ways include rejection of the data, which we have done in the training set (we removed training objects with incomplete data entries), however it could not be done on the testing set, as the LSST-DESC PZ wanted photo-z's to be estimated for every object in the given testing set. Upon inspection, we find that 2.1% of the objects in the Buzzard Gold Sample testing set have incomplete data, and most of the missing entries are from the u-band, while some have missing g and y bands. This situation led us to an interesting question: in the context of ANNZ2, how do the different methods deal with missing data and affect the overall photo-z performance of the entire sample?

To analyse this, we used 4 different methods to replace the missing entries: (1) we replace the missing entry with the mean of the band column, i.e. replace the missing u-band with 26.0 and etc, therefore the values are unique to each row; (2) we replace the missing entry with the mean of the row (i.e. $ugrizy$ magnitudes of a single galaxy), therefore the values are unique to each object; (3) we conduct a simple linear interpolation between bands, e.g. if the g band is missing, we fill it up with an average value of u and r. In the case of missing u bands, we extrapolate using g and r, and vice versa for y; and (4) we conduct a simple ANN search to input the missing value. For example, for all objects in the testing set with missing u bands, we take the training set and train with $grizy$ with the output as u, and we test on these objects; those with missing ug, we train with $rizy$ with output as ug, and etc. This could potentially be a very tedious process, however due to the dataset only missing ugy values, this was easily done without creating an automated process. Some of these methods were motivated by the discussion on replacing missing data by Batista and Monard (2003).

Table 24: Comparison of photo-z performance in σ_{RMS}, σ_{68}, η_{out}, ρ_{CRPS}, ρ_{PIT} and n_{RMS} for ANNZ2 when using different methods to deal with missing data. Also shown are the metrics when all incomplete data are removed from the sample.

Method	σ_{RMS}	σ_{68}	$\eta_{\mathrm{out}}(\%)$	ρ_{CRPS}	ρ_{PIT}	n_{RMS}
(1) Mean of column	0.0557	0.0315	2.76	0.0221	0.0303	0.150
(2) Mean of row	0.0581	0.0317	3.03	0.0223	0.0301	0.150
(3) Interpolate row	0.0537	0.0308	2.42	0.0218	0.0307	0.150
(4) Use ANN	0.0536	0.0307	2.41	0.0218	0.0307	0.151
No incomplete data	0.0538	0.0308	2.43	0.0218	0.0304	0.152

ANNZ2 was trained on all 43 491 objects from the Buzzard Gold Sample training set, the missing entries of the testing set were replaced differently according to these 4 methods, and the results of the entire Buzzard Gold Sample testing set (with both good and incomplete data) are shown in Table 24. At first glance, we notice that the σ_{RMS}, σ_{68} and η_{out} of the methods 1 and 2 are significantly higher than methods 3 and 4. This is surprising since the number of objects with incomplete data are very low: with only 2.1% of incomplete data, the difference in method can result in a difference the overall degradation of σ_{RMS} as high as 8.4%. Despite the fact that there is very little change in n_{RMS} and ρ_{PIT} across all 4 methods, this results suggest that although both are widely used missing data replacement methods, they are not suitable to be used in the context of photo-z's since each column (*ugrizy*) for each galaxy are not independent variables.

It is interesting to know how similar the results are for methods 3 and 4, and their results are essentially equal to the results when no incomplete data are in the sample at all. Figure 65 shows graphically the performance of these 4 methods, and we see that the distributions of objects with incomplete data for methods 3 and 4 are almost identical, while obvious biases are seen for methods 1 and 2. From the practical point of view, replacing the missing data using ANN actually requires much more time than just merely interpolating the data; therefore the fact that interpolation is able to achieve a result as good as an ANN is good news. Therefore, we conclude that using an interpolation between bands is the best and most practical method in solving missing data problems for machine learning photo-z algorithms.

Here we note that in general there are two kinds of missing data: *undetected* and *unobserved*, ideally they should be dealt with separately. Undetected magnitudes are magnitudes that have their fluxes observed but measured to be negative which translates to an undefined magnitudes, these are mostly faint objects with low flux values, turning negative after the sky background flux is subtracted from them. When using machine learning methods, this kind of missing data in fact could easily be dealt with by replacing the input magnitudes with fluxes, however we could not have done so here since the the Buzzard catalogue given by the LSST-PZ did not provide the fluxes, and from the catalogue we were not able to distinguish the undetected magnitudes from the unobserved magnitudes either.

Unobserved magnitudes, on the other hand, are magnitudes that have not been observed due to severe sky conditions or instrumental issues. Ideally one should not artificially "create data" to solve the problem, although to what extent is created data allowed remain a debatable topic. If we assume the strictest scenario where no data is to be created, photo-z's for objects with unobserved data should only be trained and tested on the observed data available (i.e. if only *grz*

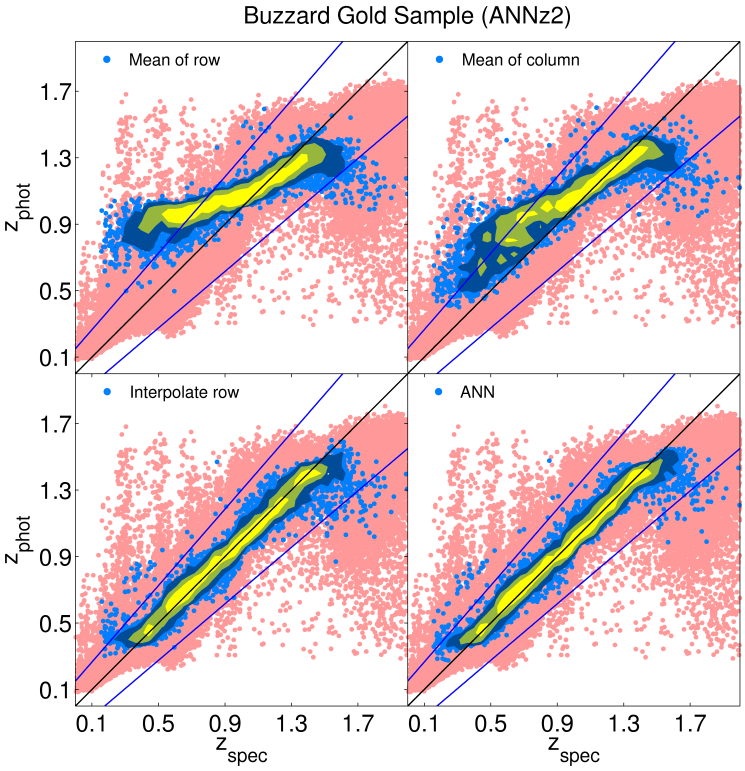

Figure 65: Photo-z v.s. spec-z by ANNz2 on the Buzzard Gold Sample, comparing the 4 different methods to deal with missing data. The red dots in the background indicate objects with complete data (results are the same in all 4 plots), while the blue dots and density contours are objects with missing data (mixture of undetected and unobserved magnitudes). The top two plots show that both imputation methods using the mean of row and column introduce high biases to the photo-z's, and thus are not feasible solutions for this work.

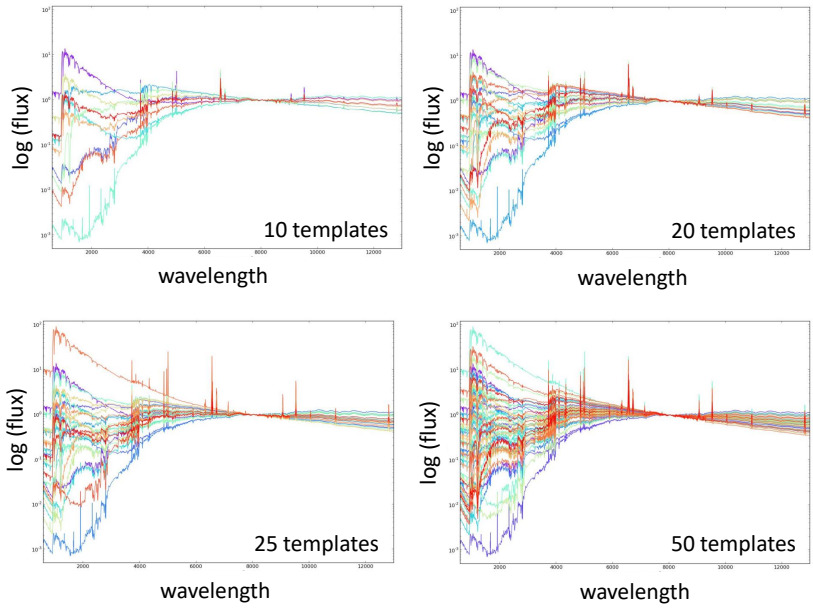

Figure 66: The ordered Buzzard SEDs, showing when 10 (top left), 20 (top right), 25 (bottom left) and 50 (bottom right) templates were used. Figures provided by LSST-DESC PZ.

bands are observed, one should train and test on *grz* bands only), which we expect to perform worse as we have established in Section 6.5.2. Therefore in this section we have only provided a quick and seemingly workable solution to the problem, however we leave the vigorous and proper analysis on this topic as a potential future work.

7.4.3 *Delight v.s. Number of Templates*

As mentioned in Section 5.6.6, the Buzzard templates are rank ordered such that a set of the first N templates will be the most spanning template set for N templates, according to the algorithm used to choose them. Figure 66 shows the Buzzard templates in different numbers respectively, to illustrate the effects of using the top N templates for photo-z estimation.

While it was recommended by LSST-DESC PZ that using 50 templates would suffice, we wanted to see the effects caused when trained with different numbers of templates for DELIGHT. Since DELIGHT uses GPs to 'correct' and produce more realistic templates based on the training set, it would be interesting to know how dependent DELIGHT is on the templates. Using 10 873 objects from the Buzzard Gold Sample training set, we estimated photo-z's using 10, 20, 25, 50, 100 and 150 ordered templates, respectively. The results of these runs are tabulated in Table 25.

From the table, we see that the performance of the point estimates still depend highly on the templates used, since the values of σ_{RMS}, σ_{68} and η_{out} are shown to decrease significantly as the number of templates increase. This is also evident from Figure 67, as it shows the improvement in scatter as the number of templates increase, although the performance levels off after 50 templates. This is true for the n_{RMS} as well, but not for ρ_{CRPS}, where there is no significant change in performance across the number of templates. This means that the number of templates have minimal impact on how well the $p(z)$ encapsulate the true redshift, which may be due to the

Table 25: Comparison of photo-z performance in $\sigma_{\rm RMS}$, σ_{68}, $\eta_{\rm out}$, $\rho_{\rm CRPS}$, $\rho_{\rm PIT}$ and $n_{\rm RMS}$ for DELIGHT when using different number of templates.

No. of Templates	$\sigma_{\rm RMS}$	σ_{68}	$\eta_{\rm out}(\%)$	$\rho_{\rm CRPS}$	$\rho_{\rm PIT}$	$n_{\rm RMS}$
10	0.1669	0.0578	11.50	0.0172	0.1465	0.166
20	0.1516	0.0410	7.43	0.0177	0.0665	0.142
25	0.1298	0.0353	5.46	0.0177	0.0524	0.140
50	0.0854	0.0292	3.61	0.0179	0.0237	0.136
100	0.0828	0.0264	3.28	0.0181	0.0315	0.133
150	0.0772	0.0259	3.14	0.0178	0.0329	0.132

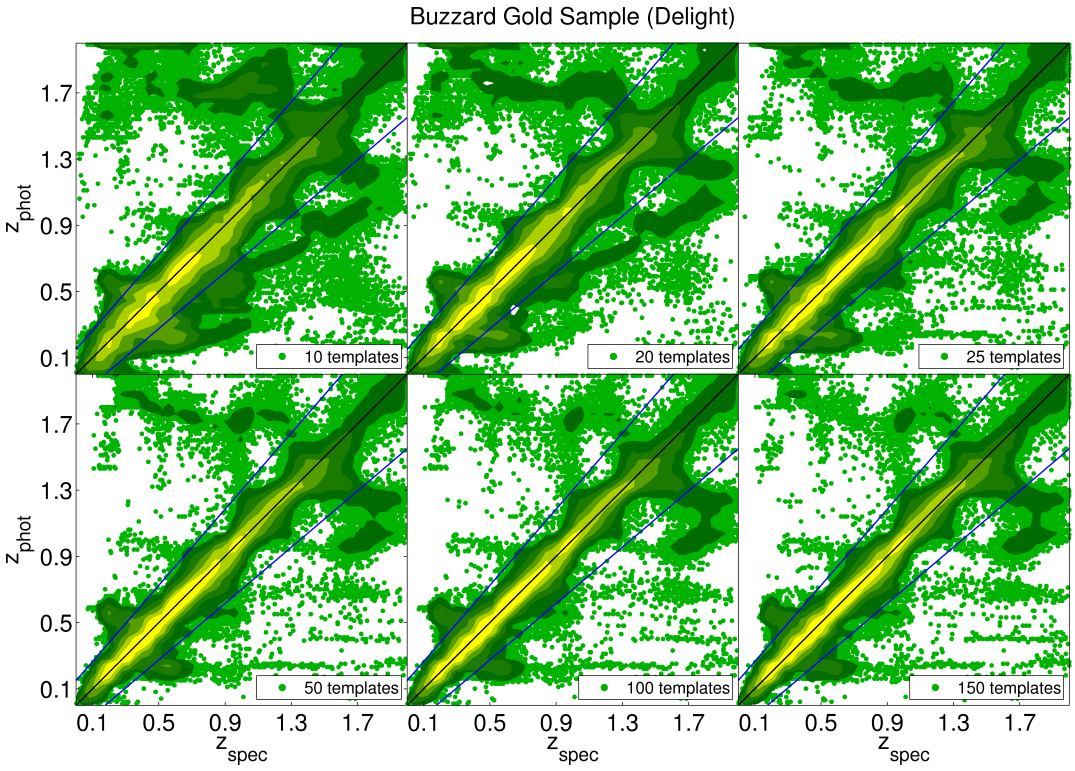

Figure 67: Photo-z v.s. spec-z by DELIGHT on the Buzzard Gold Sample, comparing the results when different numbers of the ordered Buzzard templates are used.

Table 26: Comparison of the bias, $\sigma_{\rm RMS}$, σ_{68}, $\sigma_{\rm IQR}$, $\eta_{3\sigma}$, $\eta_{\rm out}$, $\rho_{\rm CRPS}$, $\rho_{\rm PIT}$ and $n_{\rm RMS}$ between ANNz2 and DELIGHT in the context of LSST.

Code	**bias**	$\sigma_{\rm RMS}$	σ_{68}	$\sigma_{\rm IQR}$	$\eta_{3\sigma}$	$\eta_{\rm out}$	$\rho_{\rm CRPS}$	$\rho_{\rm PIT}$	$n_{\rm RMS}$
ANNz2	0.0009	0.0536	0.0307	0.0276	7.60	2.41	0.0218	0.0307	0.151
DELIGHT	−0.0024	0.0772	0.0259	0.0238	7.00	3.14	0.0178	0.0329	0.132

fact that the positions of the $p(z)$ are primarily set in the first few templates, while the addition of templates made the widths of the $p(z)$ to vary, resulting in an irregular change in $\rho_{\rm PIT}$.

Therefore, we conclude that the point estimates of DELIGHT are still heavily dependent on the templates, less so for the $n(z)$, but insignificantly for the $p(z)$ shape, given that the first N ordered templates well represent the galaxy types in the sample. Therefore, despite having the ability to create its own empirical templates, DELIGHT still relies on good templates for it to perform well.

7.4.4 *Comparison Between ANNz2 and Delight*

In this section, we take the best results from ANNz2 (method 4) and DELIGHT (150 templates), and we compare their performances in the Buzzard Sample. Although this sample is less realistic and contains a large and representative training set, this would help us understand the performance of both codes in the context of a deep 6-broadband survey like the LSST.

To compare with the analysis of the LSST-DESC PZ group in Schmidt et al. (in prep.), we introduce 3 metrics used in that work which have not been used in any part of this thesis. The first is the *bias*, which is defined to be the median of the distribution of $\Delta z_i = \frac{z_{{\rm phot},i} - z_{{\rm spec},i}}{1+z_{{\rm spec},i}}$. The median was chosen instead of the mean in this case which is less sensitive to outliers. The second metric is $\sigma_{\rm IQR}$, defined as

$$\sigma_{\rm IQR} = \frac{\rm IQR(\Delta z_i)}{1.349} \tag{116}$$

where IQR is the *interquartile range* (difference between the 75th and 25th percentile), divided by 1.349 as there is a factor difference between the IQR and the standard deviation of a normal distribution. This metric is used in place of $\sigma_{\rm RMS}$ to describe the scatter in Schmidt et al. (in prep.) as it is less susceptible to outliers. Finally the outlier fraction is described by $\eta_{3\sigma}$, the $3\sigma_{\rm IQR}$ outliers, i.e. the fraction of objects with $|\Delta z_i| > \max(0.06, 3\sigma_{\rm IQR})$, and this is required to be less than 10% according to the photo-z requirement of LSST. These and other metrics used for ANNz2 and DELIGHT are tabulated in Table 26.

We first evaluate their performances with regard to the point estimates. From the values of $\sigma_{\rm RMS}$ and $\eta_{\rm out}$, we see that ANNz2 performs better than DELIGHT, however for the σ_{68} DELIGHT does better. This result is typical when comparing between template-based and machine learning methods, since training based methods result in getting an 'average' relationship between the magnitudes and redshift, giving rise to the higher scatter.

Figure 68 visualises the point estimate performance between both codes, we note that we are comparing the best point estimate values from both codes, which is the mean of the $p(z)$ for ANNz2, and the peak of the $p(z)$ for DELIGHT. Other than the difference in scatter and σ_{68} as

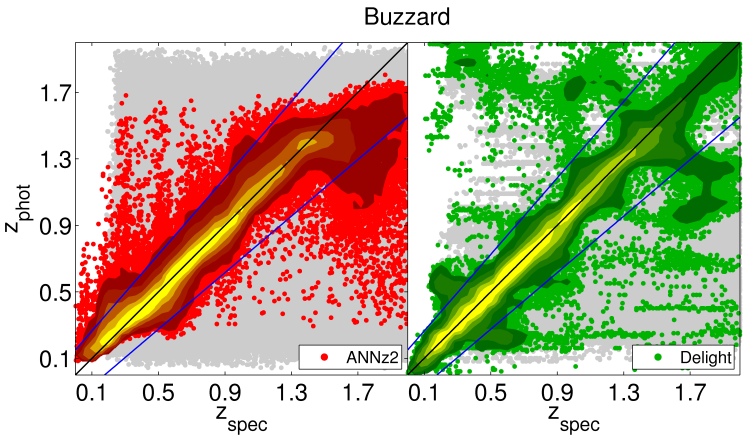

Figure 68: Photo-z v.s. spec-z by ANNz2 and DELIGHT on the Buzzard Sample, with the coloured dots referring to objects in the Gold Subsample. Note that training was done on gold sample objects only ($i < 25.3$), and tested on all objects in the sample.

described earlier, we see that DELIGHT performs significantly better at higher redshift, although it has many catastrophic outliers at high $z_{\rm phot}$. ANNz2, on the other hand, suffers from biases at the upper and lower end of redshifts, this is due to the fact that machine learning algorithms would focus on training objects which are better represented in the sample, therefore getting better results at intermediate redshifts.

From Figure 68 we can also see the photo-z's obtained for the non-gold sample objects (grey). ANNz2 was made to train on only objects with $i < 25.3$ in order to maximise the results for the Gold Sample, however this causes degradation of results for objects with $i > 25.3$, where the photo-z's are literally everywhere in the background. The photo-z's for these objects can improve by conducting a separate training for them. We see better performance in DELIGHT in this respect, however with photo-z's having a higher tendency to be underestimated. However we note that most science cases will be studied using the Gold Sample, therefore bespoke optimisations for objects fainter than the Gold Sample can be done as required.

Thus far we have found that both ANNz2 and DELIGHT have achieved the LSST photo-z requirement for both the bias and $\eta_{3\sigma}$, however the root-mean-square scatter $\sigma_{\rm RMS}$ for both codes remains above 0.05. The more optimistic measurement of $\sigma_{\rm IQR}$ is shown to be below 0.05, but has not reached the more stringent goal of 0.02. We decided to systematically cut the sample using odds cuts on the photo-z's from both codes to see if the requirements of $\sigma_{\rm RMS}$ would be achieved, and this is visualised in Figure 69. From the figure, we see a contrast between the performance of ANNz2 and DELIGHT. In the case of ANNz2, with a 90% odds cut, the scatter has successfully dropped below the targeted 0.05, however it only hits the 0.02 goal at 70% odds cut for objects with $i < 23.5$, and at 40% cut for objects with $i < 24.5$. DELIGHT on the other hand, although it performs relatively better for brighter objects, the odds cuts do not help to reduce the scatter below 0.05 for objects with $i < 24.5$, or reach the goal of 0.02 at all. This means that there are a lot of false positives in the sample: there are many objects in the sample which have $p(z)$ with very narrow peaks yet not encapsulating the $z_{\rm spec}$. We note however that this effect may have been caused by the catastrophic outliers, although small in numbers but giving a large scatter

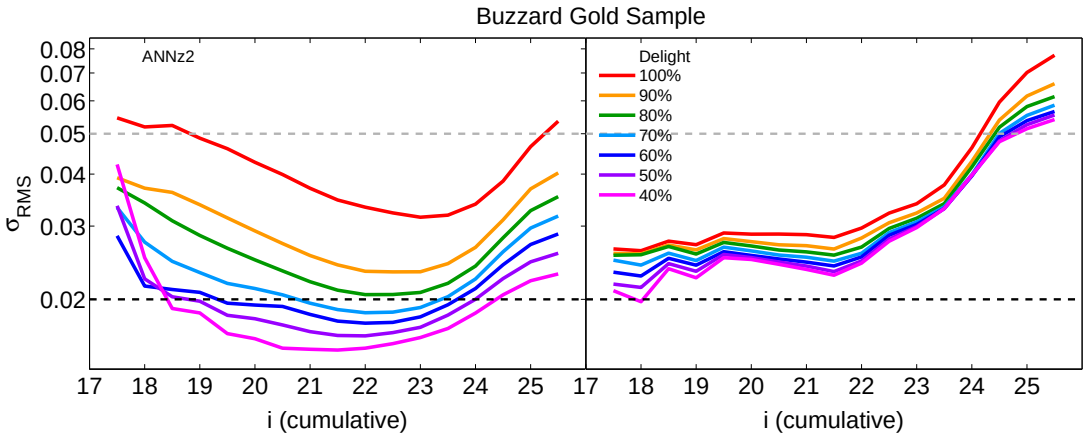

Figure 69: $\sigma_{\rm RMS}$ of photo-z's for objects brighter than i for the Buzzard Gold Sample produced by ANNz2 (left) and DELIGHT (right), when the sample is cut in the odds parameter (100% to 40%).

to the results. We expect DELIGHT to perform better in this regard if the σ_{68} was used instead (see Figure 87). Therefore we conclude that ANNz2 is very close to achieving the goal of LSST with the help of odds cuts, however we note that such cuts may introduce selection effects which may affect the science results, and thus should be only be done with extra caution. Further investigation or optimisation on DELIGHT hyperparameters are needed for it to meet the expected requirements.

We now turn our focus to compare the quality of $p(z)$ produced by both codes. From Table 26, the values of $\rho_{\rm CRPS}$ show that the $p(z)$ of DELIGHT encapsulate the $z_{\rm spec}$ better than ANNz2, while $\rho_{\rm PIT}$ shows little difference between ANNz2 and DELIGHT, indicating that both produced $p(z)$ with average widths overall. Figure 70 shows a sample of 12 $p(z)$ produced by both ANNz2 (red) and DELIGHT (green), and we see the familiar signature fuzzy ANNz2 $p(z)$ in comparison with the smoother $p(z)$ by DELIGHT. Here we can see that the narrower $p(z)$ of DELIGHT result in a smaller $\rho_{\rm CRPS}$. We also see interesting dual peaks in DELIGHT especially for higher redshifts, and in most cases at least one of the peaks has predicted the spec-z correctly. In most instances the peaks of ANNz2 coincide with DELIGHT, showing good agreement between both codes.

Figure 71 shows the QQ plot for both codes, while Figure 72 shows the distribution of PIT values. From the QQ plot we find that both codes produce relatively good photo-z PDFs on average, although ANNz2 performs slightly better in this area. From the QQ plot we see that ANNz2 on average produces $p(z)$ which are slightly wider than usual ('Z' shape curve), and this is also seen in the PIT distribution with a peak at 0.5. DELIGHT on the other hand, seems to underpredict the $z_{\rm spec}$ (curve above diagonal), meaning that most of the $p(z)$ peaks are to the left of $z_{\rm spec}$. This is also verified in Figure 68, where we see an overdensity of objects in an area slightly below the diagonal at $z_{\rm spec} \approx 1.7$.

The PIT distribution, although related to the QQ plot, provides extra information by telling us the number of outlier $p(z)$, whereby the $z_{\rm spec}$ is totally outside of the range of the $p(z)$. This can be seen from the spikes at both ends of the distribution in Figure 72. From the plot we see that there are more catastrophic $p(z)$ in DELIGHT than in ANNz2, especially at the higher end, i.e. these objects have a $p(z)$ lower than the $z_{\rm spec}$. Once again from Figure 68 we believe these objects

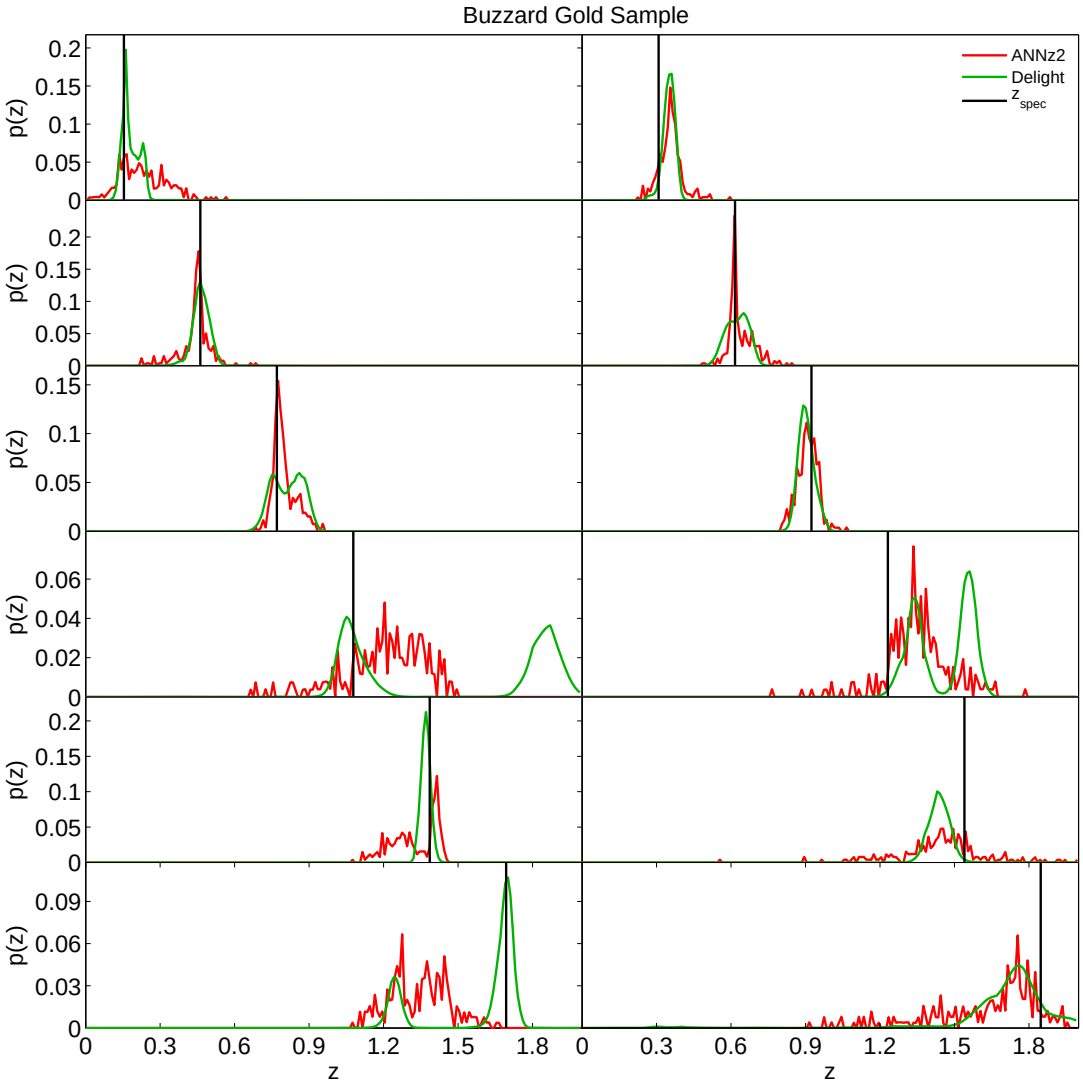

Figure 70: Sample of 12 $p(z)$'s produced by ANNz2 (red) and DELIGHT (green) for the Buzzard Gold Sample, at different true redshifts (black).

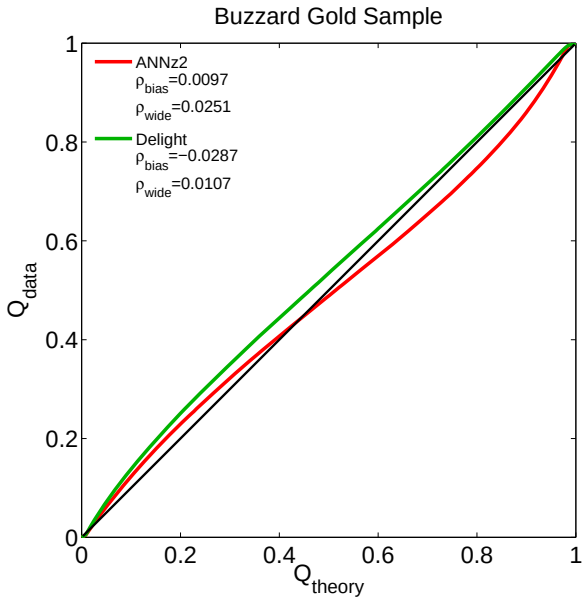

Figure 71: QQ plot comparing the $p(z)$ quality for the Buzzard Gold Sample produced by ANNz2 (red) and DELIGHT (green), respectively.

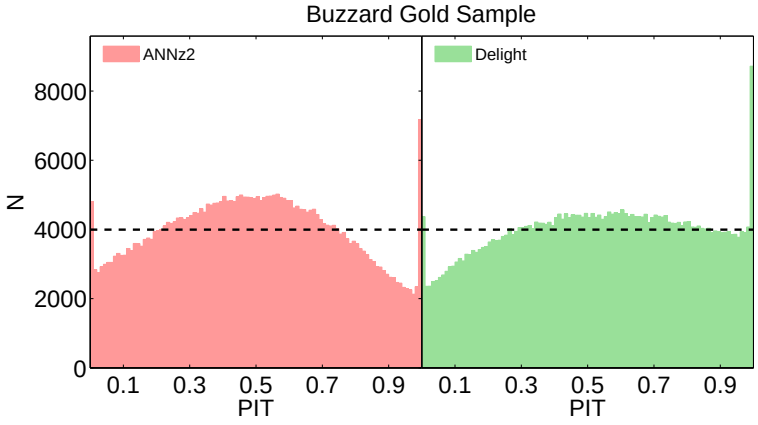

Figure 72: Distribution of PIT values comparing the performances between ANNz2 (red) and DELIGHT (green) on the Buzzard Gold Sample.

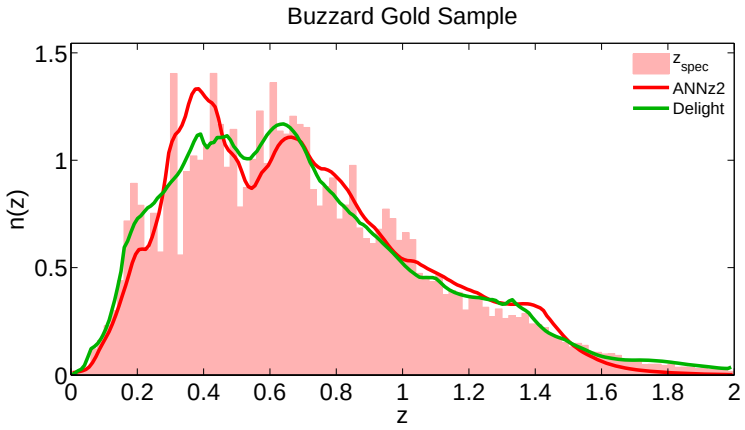

Figure 73: Redshift distribution of the Buzzard Gold Sample, comparing the distribution of the true redshift (pink histogram) with the photo-z distributions (stacked PDFs) of ANNz2 (red) and DELIGHT (green).

belong to the scatter below the diagonal of the plot, which we suspect has to do with issues of the template set given. This can be verified further if the results are rerun using a different set of templates (e.g. CWW or Brown) to see if these scatter still remain. It might however also be due to degeneracies in colour space at $z_{\text{spec}} > 1.5$ associated with the position of the Lyman break in the filters.

Finally we compare the $n(z)$ produced by the stacked $p(z)$ from each of the codes, as shown in Figure 73. From the plot we see that in fact both codes perform really well in representing the $n(z)$ of the Gold Sample. DELIGHT arguably performs better, as it correctly represents the shoulder at $z < 0.2$, and ANNz2 is shown to underpredict the higher redshift tail, which is a generic result of the machine learning algorithm as mentioned earlier.

We summarize the comparison between ANNz2 and DELIGHT as follows. ANNz2 performs better in scatter and in the overall $p(z)$ shape, it has fewer catastrophic $p(z)$, and with proper odds cuts would be able to reach the photo-z requirement target of $\sigma_{\text{RMS}} = 0.05(1+z)$ by LSST, and close to the $0.02(1+z)$ goal if using a stronger odds cut. DELIGHT, on the other hand, performs better in σ_{68}, its $p(z)$ encapsulating z_{spec} better, and produces a better $n(z)$ for the overall Gold Sample. We conclude that both codes have their advantages, and should be used in conjunction with one another, and selected based on the needs of the science cases. More effort could be put to further rectify the issues that both codes face, or even find ways to combine both results, however that will be left as a future work.

7.5 CONCLUSION

In this chapter we have studied two important aspects of the photo-z PDFs and redshift distributions: the impact of morphology on them, and the viability of various metrics to quantify and characterise them. From the first part of this chapter, we find that galaxy morphology brings improvement to the $p(z)$ widths, odds parameter, and how well the $p(z)$ encapsulates the true redshift. Galaxy morphology is found to have little impact on $n(z)$, however a relatively good

$n(z)$ can be constructed using only 2 broadbands and morphology. This information has been applied to the CS82 Morphology Catalogue, where photo-z's, morpho-z's and redshift PDFs for galaxies as deep as $i \sim 24$ have been produced.

The performance of ANNZ2 and DELIGHT using a representative training set in the context of LSST ($i \sim 25.3$) has also been tested, and we find that both codes are competitive in their own ways, and that ANNZ2 would be close to achieve the most stringent photo-z requirements of LSST at an odds cut of 40%. In the process, we find that the KS statistic is not a suitable metric to evaluate $n(z)$, since the spectroscopic training sets, whether weighted or unweighted, are usually not smooth enough to for this test to work. The CRPS was found to be good to characterise the ability of the $p(z)$ to forecast the spec-z.

ρ_{PIT} however only tells us on average if the $p(z)$ widths are well distributed, within the metric itself we do not know if the $p(z)$ are too wide / narrow, or over- / under-predicted. While the QQ plots, PIT histograms, ρ_{bias} and ρ_{wide} provide this information, it remains a question as to how this information can be put to good use, e.g. if the point estimates are very accurate to begin with, does it still matter if the $p(z)$ are too wide or too narrow? It is yet to be understood how ρ_{PIT} has an impact on n_{RMS}, and along with the large number of unexplored performance metrics, these could remain as work to be explored in the future.

Table 27: List of headers and their respective descriptions for the CS82 Photo-z Catalogue, which could be found at ftp://ftp.star.ucl.ac.uk/johnsyh/cs82/. The file cs82_phz.fits has all information below except for the $p(z)$ (last row). The files cs82_pz_phot.fits and cs82_pz_morph.fits contain the $p(z)$ for each object, based on the training with $ugriz + \mu rqnp$ and $i + \mu rqnp$, respectively. Values that are not available are left blank in the catalogue.

Column Name	Description
OBJID_CS82	CS82 object ID
RA	right ascension (deg)
DEC	declination (deg)
FLAGS	source extraction quality flag (ranges from 0 to 3, with 0 being the best)
WEIGHT	LENSFIT shape measurement flag. WEIGHT>0 indicates good shape measurement
FITCLASS	LENSFIT star-galaxy classifier, with values 1 indicating stars, 0 for galaxies, -1 for no usable data, -2 for blended objects, -3 for miscellaneous reasons and -4 when the χ^2 exceeded the critical value
MAG_AUTO	CS82 Kron i-band magnitude
MAGERR_AUTO	CS82 Kron i-band magnitude error. Signal-to-noise ratio is 1.086 divided by MAGERR_AUTO
MAG_EXP	CS82 exponential fit i-band magnitude
MAG_PSF	CS82 PSF i-band magnitude
REFF_EXP	exponential fit effective radius (arcsec)
ASPECT_EXP	exponential fit axis-ratio
MU_MEAN_EXP	exponential fit mean surface brightness
P_EXP	exponential shape probability, with values close to 1 indicating that the object has a higher probability being a disc galaxy, and an elliptical galaxy otherwise
N_SER	Sérsic index
SPREAD_MODEL_SER	Sérsic spread model, the star-galaxy separator we used for this study. All objects in this catalogue have SPREAD_MODEL_SER>0.008, which are considered extended objects (galaxies)
LENS	lensing tag. Objects with LENS = 1 are objects from the lensing subsample (FITCLASS>0, WEIGHT>0)
OBJID_SDSS	SDSS object ID for objects with matched $ugriz$ broadband magnitudes (if available)
MAG_DERED_U	SDSS dereddened u-band magnitude (if available)
MAG_DERED_G	SDSS dereddened g-band magnitude (if available)
MAG_DERED_R	SDSS dereddened r-band magnitude (if available)
MAG_DERED_I	SDSS dereddened i-band magnitude (if available)
MAG_DERED_Z	SDSS dereddened z-band magnitude (if available)
ZSPEC	spectroscopic redshift (if available)
ZPHOT	photometric redshift estimated using inputs $ugriz$+morphology (if available)
ZMORPH	photometric redshift estimated using inputs i+morphology

Column Name	Description
ZBEST	best redshift for this object, in order of priority: ZSPEC, ZPHOT, ZMORPH
ODDS_PHOT	odds value for ZPHOT
ODDS_MORPH	odds value for ZMORPH
ODDS_BEST	odds value for ZBEST (= 1 if ZSPEC is used)
SOURCE_SPEC	source of spectroscopy: SDSS, BOSS, DEEP2, WIGGLEZ or VVDS (if available)
CLASS_SPEC	class of object based on spectral fit: GALAXY or QSO (if available)
Z_0 - Z_189	$p(z)$ values for ZPHOT / ZMORPH, ranging from $z = 0.005$ to $z = 1.895$ in an equal step size of 0.01

8

PROJECT 3: PHOTOMETRIC REDSHIFTS ON THE PAU NARROWBAND SURVEY

Enter through the narrow gate. For wide is the gate and broad is the road that leads to destruction, and many enter through it.

Matthew 7:13

8.1 INTRODUCTION

In Chapter 6 and Chapter 7 we have conducted photo-z estimation using broadband filters, and thus far we have seen that machine learning algorithms perform well for broadband surveys. However, as the number of filters increases by a large number, we expect machine learning codes to perform less effectively due to the curse of dimensionality. Machine learning codes have to be optimised with more complex architectures to get better results in this situation, since increased dimensionality of the input space would increase the likelihood of converging to a local rather than the global minimum, therefore require larger training sets. In the special case of ANNs, the curse of dimensionality is exemplified due to the full connectivity between nodes, and also its insensitivity to spatial structure data (e.g. the relationship between adjacent band filters are treated the same as as those far apart). A more suitable substitute for ANN in this scenario is the *convoluted neural network*, which however is beyond the scope of this work.

In fact photo-z estimation on large numbers of bands is not new. Recent photo-z estimation work which uses many bands include the CFHTLS (Ilbert et al. 2006) which used 11 bands ($ugriz$ from CFHT and $BVRIJK$ from VVDS); COSMOS (Ilbert et al. 2009) used 30 bands (from CFHT, Subaru, UKIRT, Spitzer, GALEX and the National Optical Astronomy Observatory (NOAO)) with mixed bandwidths and wavelengths ranging from the ultraviolet to mid-infrared; the Advance Large Homogeneous Area Medium-Band Redshift Astronomical (ALHAMBRA, Molino et al. 2014) Survey used 23 bands (20 medium optical bands and JHK); and most recently, KiDS (Bilicki et al. 2017) used 12 bands ($ugri$ from KiDS, $zYJHK$ from the VISTA Kilo-degree Infrared Galaxy (VIKING, Edge et al. 2013) Survey and $W_1W_2W_3$ from the Wide-field Infrared Survey Explorer (WISE, Wright et al. 2010). However this work is different from those mentioned as we are considering a large number of only narrow bands with no overlaps. In this situation emission line features are exceptionally significant, and it would be interesting to see if machine learning methods are able to capture this information and produce accurate photo-z's for them. Besides, with a large number of narrow bands in a single survey, the probability of missing or undetected

data becomes prominent, and machine learning methods may need to have a workaround for such situations.

In this chapter we move our analysis from broadband surveys to a narrowband survey. We have indeed a perfect example for this analysis: the PAU Survey (see Section 4.1.5 for full description), which conducts observations in 40 narrow bands with wavelengths ranging from 455 nm to 855 nm. We will evaluate the performances of the 2 machine learning codes used previously (ANNZ and ANNZ2) with the template-based code BPZ and the hybrid code DELIGHT. The focus will be on DELIGHT: we will optimise the algorithm over different settings, template sets and calibrations to obtain competitive photo-z's for the PAU early data release in comparison with the advanced template code BCNZV2 (Eriksen et al., in prep.). The goal we have is to achieve a photo-z performance of $\sigma_{68} < 0.0035(1+z)$, which is the target requirement of the PAU Survey itself (Martí et al. 2014).

This chapter is divided mainly into three sections. In Section 8.2.1 and Section 8.2.2 we work on PAU simulated samples: in the former we compare the performance of different machine learning algorithms of ANNZ2 with DELIGHT, while in the latter we study and select the most suitable template to be used. And finally in Section 8.3 we move to real data: the early data release of PAU, where we calibrate the fluxes of broadbands to be trained in conjunction with the narrow bands, and show the final photo-z performance for DELIGHT compared to other codes. The calibration process is discussed and analysed, and the performance of DELIGHT with missing or incomplete data is studied too. The future prospects of DELIGHT on narrowband surveys is also discussed.

8.2 PHOTOMETRIC REDSHIFTS ON SIMULATED GALAXY SAMPLES

8.2.1 *PAU Mock Galaxy Sample*

To start our photo-z analysis on on the PAU narrowband survey, we use the PAU Mock Galaxy Catalogue constructed by Martí et al. (2014) as described in section Section 4.2.2. This set was chosen as it is a relatively simple and clean sample where we could not only test the performance of machine learning algorithms of ANNZ2, but also use it to optimise the initial settings of DELIGHT. In fact the public code DELIGHT underwent major updates as a result of this work, the current updated DELIGHT code (as of May 2018) can now process flux likelihoods in a numerically stable manner for a large number of narrowbands.

The mock catalogue contains 244 754 objects with 6 *ugrizy* broadbands and 40 narrow bands (filter responses shown in Figure 13 and Figure 14 respectively), each object having a magnitude $i < 22.5$ and redshift $z_{\rm spec} < 2.0$. We selected a subset of 40 792 objects as our testing set, and the distributions of i and $z_{\rm spec}$ are shown in Figure 74.

In this analysis both the performances of ANNZ and ANNZ2 are compared, each code is trained and validated with 40 792 objects. For ANNZ2, we specifically set the code such that it produces photo-z's with only one machine learning algorithm at a time, allowing us to compare the performance of ANNs, BDTs and KNNs separately. Each run in ANNZ2 uses 5 committees with different random seed or settings so that the final results is a weighted average of it. An extra

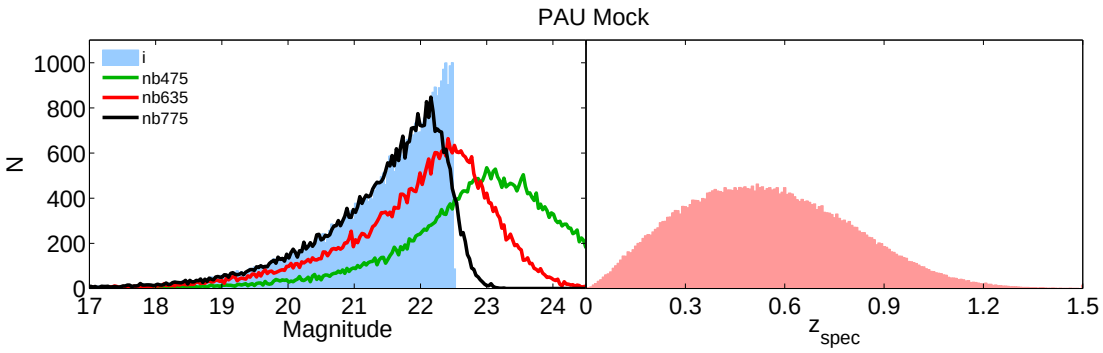

Figure 74: Distributions of the *i*-band (left) and $z_{\rm spec}$ (right) for the PAU Mock Galaxy Catalogue by Martí et al. (2014). Also shown are the distributions of the narrowband magnitudes *nb*475, *nb*635 and *nb*775 respectively.

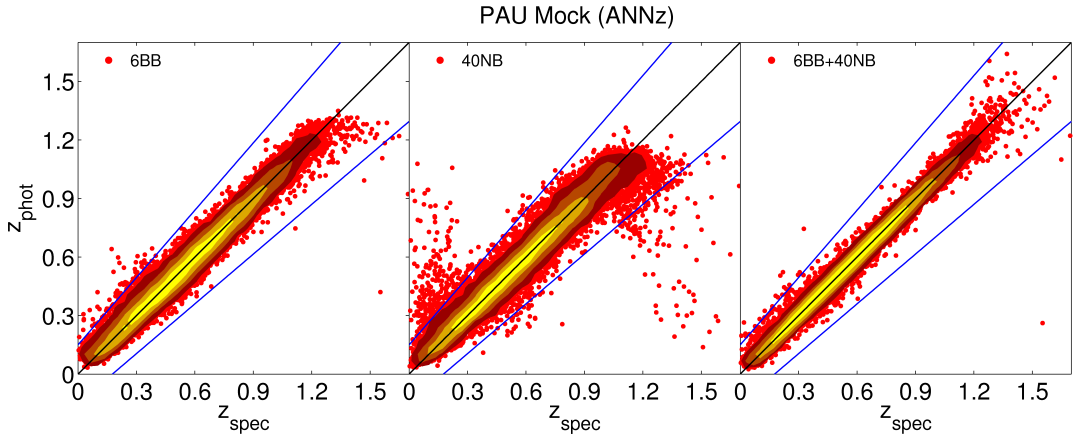

Figure 75: Photo-z v.s. true redshift (denoted as $z_{\rm spec}$ throughout this chapter) performance of ANNz, when trained with 6 bands (left), 40 bands (middle) and 46 bands (right), respectively, on galaxies from the PAU Mock Galaxy Catalogue.

run with mixed machine learning algorithms was also conducted, this run uses 5 committees using 2 runs each with ANNs and BDTs and 1 run with KNN.

As for DELIGHT, the default settings were used, however the code was trained using only 10 198 objects due to the limitations of project time. Note however that the training process is fundamentally different in DELIGHT, so that a small training set is unlikely to be a disadvantage. We also note that, as the PAU Mock Galaxy Catalogue was generated with older versions of filter response curves, we revert to use the same curves for our training and testing for this sample as well. With each code and setting, 3 separate trainings were conducted: the first trained with only the 6 PAU broadbands, the second with only the 40 narrowbands, and the third with all 46 bands. The performances for each run are summarised in Table 28.

From Table 28 we see an overall trend from all codes that the scatter and outlier fraction are at their maximum when trained with only 40 narrow bands, and minimum when trained with 46 bands. Figure 75 visualises the results of ANNz, where we can see visible scatter when trained using 40 bands only: a visible portion of objects at very low and very high $z_{\rm spec}$ are set to have

Table 28: Photo-z performances of ANNZ and ANNZ2 for the PAU Mock Galaxy Catalogue, shown with the respective number of training bands n_b, and the performance metrics $\sigma_{\rm RMS}$, σ_{68}, and $\eta_{\rm out}$ respectively.

Code	n_b	$\sigma_{\rm RMS}$	σ_{68}	$\eta_{\rm out}$ (%)
ANNZ	6	0.0299	0.0250	0.17
ANNZ2 ANN	6	0.0335	0.0280	0.27
ANNZ2 BDT	6	0.0301	0.0240	0.25
ANNZ2 KNN	6	0.0339	0.0280	0.29
ANNZ2 mixed	6	0.0295	0.0238	0.18
ANNZ	40	0.0441	0.0256	1.05
ANNZ2 ANN	40	0.0457	0.0269	0.99
ANNZ2 BDT	40	0.0474	0.0258	1.44
ANNZ2 KNN	40	0.0527	0.0328	1.98
ANNZ2 mixed	40	0.0444	0.0247	1.18
ANNZ	46	0.0179	0.0124	0.06
ANNZ2 ANN	46	0.0196	0.0146	0.06
ANNZ2 BDT	46	0.0279	0.0172	0.27
ANNZ2 KNN	46	0.0318	0.0205	0.40
ANNZ2 mixed	46	0.0214	0.0144	0.10

photo-z's all across the range. The fact that the results of 40 bands seem to under-perform when compared to 6 bands is due to the higher signal-to-noise of the broadbands when compared to the narrowbands. Indirectly, we also see that the machine learning algorithms are more sensitive to bigger and broader features (broadbands) than the narrow bands: the wider wavelength range and higher signal-to-noise of the *ugrizy* bands seem to reduce the overall scatter and gets the extreme low and high redshifts right, while the narrowbands helped to fix the σ_{68}, and the combination of both provides the best possible performance.

Comparing the results of different machine learning algorithms, we find that relatively BDTs perform the best when the number of inputs are small, but at large number of inputs ANNs performs the best. KNNs perform the worst across all three cases, the best result is obtained when a mixture of machine learning algorithms is used, bringing up to 20% improvement in σ_{68} in certain cases. This is so because since each machine learning algorithm responds to different parts of the data differently, the combined photo-z PDF weighted according to each code output provides more accurate photo-z than using them individually. Despite that, when the results are inspected visually (Figure 76), the difference is actually not very obvious. In terms of code computing speed, ANNs take about 10 times longer to run than to BDTs and KNNs, but we recommend a mixture of machine learning algorithms, provided that the hyperparameters of each algorithm are well optimised.

Despite performing relatively well, the performance of machine learning algorithms is still far from reaching $\sigma_{68} < 0.0035(1+z)$, therefore we move towards the use of DELIGHT. As we have described in Section 5.5, DELIGHT combines the use of template and machine learning, and we have seen its performance on 6 LSST broadbands in Section 7.4. As part of this project, DELIGHT was updated in November 2017 and later April 2018 after bugs were found which prevented the code from estimating photo-z's for 40 narrowbands. The above exercise for ANNZ and ANNZ2 is

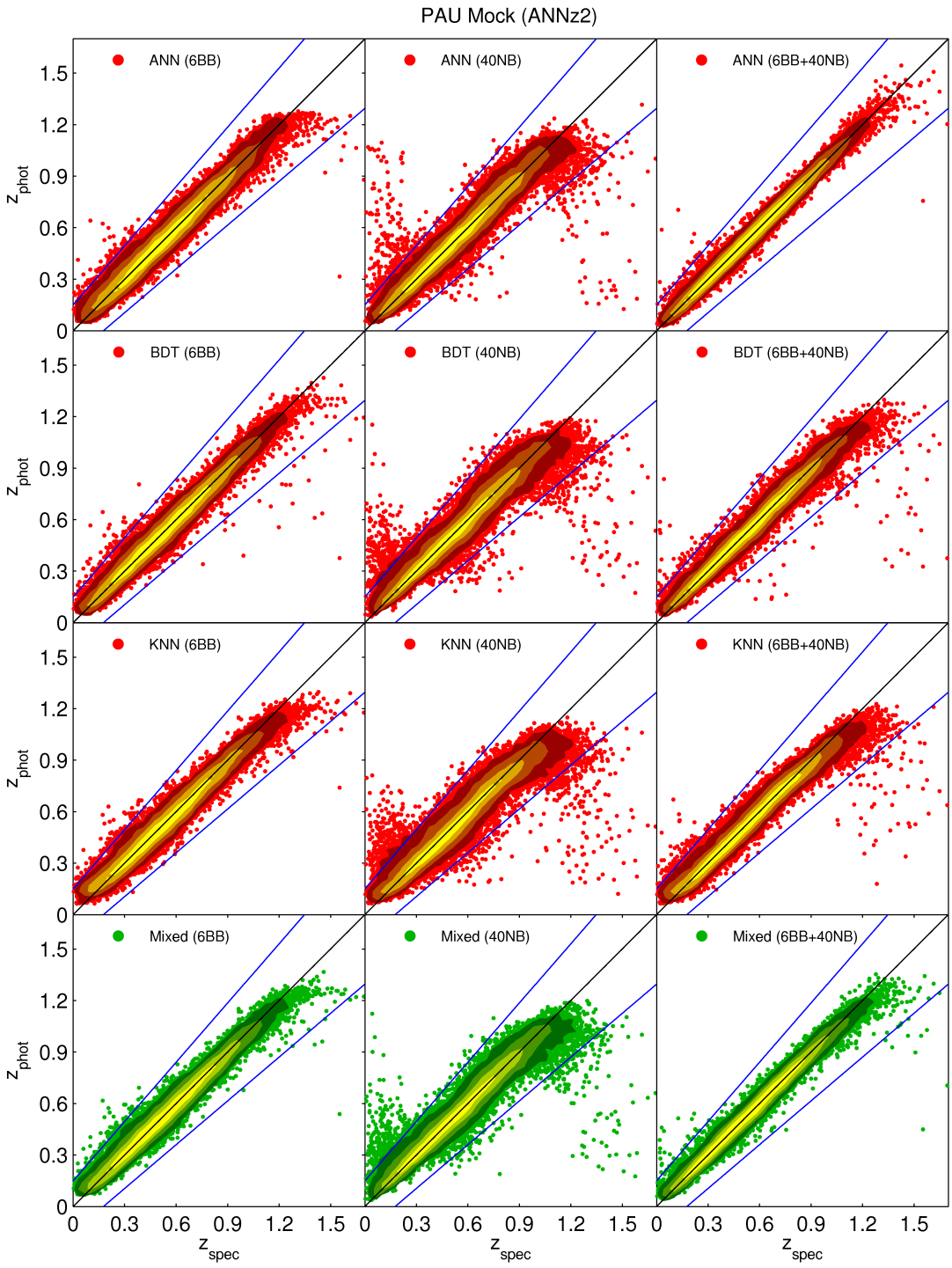

Figure 76: Photo-z v.s. spec-z performance of ANNz2, when trained with 6 bands (left column), 40 bands (middle column) and 46 bands (right column), respectively, on galaxies from the PAU Mock Galaxy Catalogue. The rows from top to bottom show the use of different machine learning algorithms, namely ANN, BDT, KNN and a mixture of all three respectively.

Table 29: Photo-z performances for DELIGHT on the PAU Mock Galaxy Catalogue, shown with the respective number of training bands n_b, and the performance metrics σ_{RMS}, σ_{68}, η_{out}, and their relative percentage improvements $\Delta\%$ respectively.

Code	n_b	σ_{RMS}	$\Delta\%$	σ_{68}	$\Delta\%$	η_{out} (%)	$\Delta\%$
Template	6	0.0291		0.0211		0.32	
GP	6	0.0281	3.5	0.0207	1.8	0.28	11.5
Template	40	0.0667		0.0057		1.90	
GP	40	0.0423	36.5	0.0086	−51.6	1.27	32.8
Template	46	0.0153		0.0049		0.09	
GP	46	0.0154	−0.5	0.0069	−39.4	0.08	16.2

now repeated for DELIGHT using its own template code and GP code. The results are summarised in Table 29. In this section the LEPHARE SEDs were used in DELIGHT; this same template set was also used to generate this mock galaxy catalogue.

The performance of DELIGHT on the 6 broadbands is comparable to both ANNZ and ANNZ2, however we see a big jump in performance in the σ_{68} when trained with 40 and 46 bands, reaching $\sigma_{68} = 0.005(1+z)$, this is visibly significant in Figure 77 and not achievable by any pure machine learning method. We also notice the distinct difference between DELIGHT and both ANNZ and ANNZ2: with the help of SED templates, the training of DELIGHT is more sensitive to small features of the change in flux, resulting in improved results in σ_{68} when trained with 40 bands as compared to 6 broadbands.

We also compare the improvement between the template code and the GP code of DELIGHT. For training with 6 or 40 bands, we see mostly improvement in the GP code in terms of scatter and outlier fraction when compared to the template results, justifying that with the help of the training set, it is able to refine the existing LEPHARE templates to give better results. However we see degradation in σ_{68} for 40 of 46 bands when moving from template to GP, we argue that this is a result of training uncertainties and the high number of inputs, mostly due to the fact that the values of σ_{68} here are very low to start off with (at least 2 to 4 times better than those trained in broadbands), the change is supposedly small but perceived big when viewed in percentages. In fact this degradation is not at all visible in Figure 77, we have also checked results from the following sections and find improvement in σ_{68} from template to GP in all cases. Ultimately these results are promising, showing that DELIGHT might be capable of improving the template photo-z estimations in both narrow and broadband surveys, mostly emphasising on reducing the scatter and outlier fraction.

Here we conclude that DELIGHT is definitely the more suitable code to use for narrowband surveys as compared to ANNZ and ANNZ2. This result supports the conclusion of Alonso et al. (2017) who also found that the performance of ANNZ on narrowbands is not on par with that of BPZ, a template-based code. DELIGHT provides the best of both worlds: it keeps the low σ_{68} of template based codes, while reducing the scatter with the help of the machine learning algorithm GP. Therefore in the next section we will continue to optimise DELIGHT for this purpose.

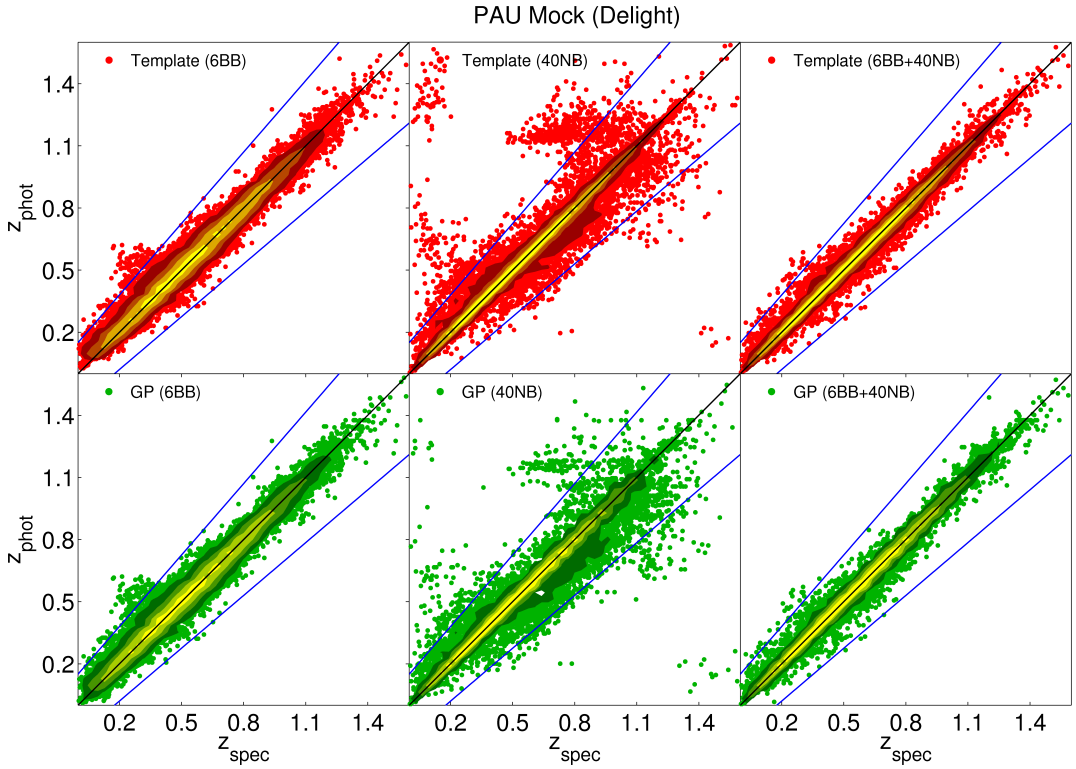

Figure 77: Photo-z v.s. spec-z performance of DELIGHT, when trained with 6 bands (left column), 40 bands (middle column) and 46 bands (right column), respectively, on galaxies from the PAU Mock Galaxy Catalogue. The top row shows the results of the DELIGHT template code, and the bottom row the DELIGHT GP code.

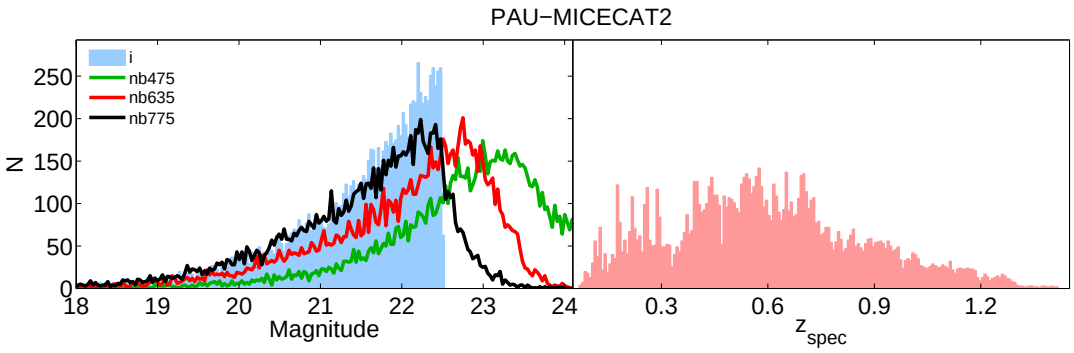

Figure 78: Distribution of i, $nb475$, $nb635$, $nb775$ (left) and z_{spec} (right) for the PAU-MICE2 Simulated Sample.

8.2.2 *PAU-MICE2 Simulated Sample*

Moving on from the PAU Mock Galaxy Catalogue, we now try the performance of DELIGHT on a different simulated sample. The PAU-MICE2 Simulated Sample (or PAU-MICECAT2) is simulated from the MICE N-body simulation, with the PAU narrowband filters generated using COSMOS SEDs (see Section 4.2.3 for full description of the sample). In this sample we will test the performance of DELIGHT when using different SED templates for modelling.

The original catalogue has 1 071 361 objects, we used about 1% of this catalogue as our testing set, cutting in magnitude ($i < 22.5$) and redshift ($0.05 < z_{spec} < 1.45$) yielding a sample of 10 721 test galaxies. The distribution of i and z_{spec} of this sample is shown in Figure 78. Version 2 of this catalogue was used; in this version the narrowband fluxes are without noise and emission lines, but Gaussian flux errors were generated for each object. The broadband fluxes are left out from this analysis due to some issues with simulating the S/N of the broadbands, and as of the time of writing the catalogue is still being updated. Therefore all runs below only use the 40 narrowbands as input.

In this sample we train DELIGHT with 10 721 galaxies using 5 different SED templates, namely the BPZ templates, CWW-Kinney templates, LEPHARE templates, COSMOS templates and the Brown templates (see Section 5.6 for a brief description of these templates). The similarities and differences of these templates and their impacts on the photo-z's will be discussed below as well. Using the same training set we produce photo-z's using ANNZ and ANNZ2 too for comparison, photo-z's from BPZ was also produced using the COSMOS prior (Laigle et al. 2016) and Brown templates. The results are summarised in Table 30 and graphically presented in Figure 79.

The different template runs show very interesting results. Qualitatively, from the looks of Figure 79 the results for BPZ and CWW-Kinney templates are very similar as expected, since both templates come from almost the same source, with BPZ being a little better with the help of the two bluer SSP templates it has from Bruzual and Charlot (2003). However, both templates suffer the same island of galaxies slight upper-left of the diagonal. In the run using LEPHARE templates, we see the island less prominent, with visibly less scatter from many regions and a lower σ_{68} agreeing with the results of Arnouts et al. (2002). Since LEPHARE templates were constructed by mainly interpolating the 4 CWW templates (spectral types Ell/So, Sbc, Scd and

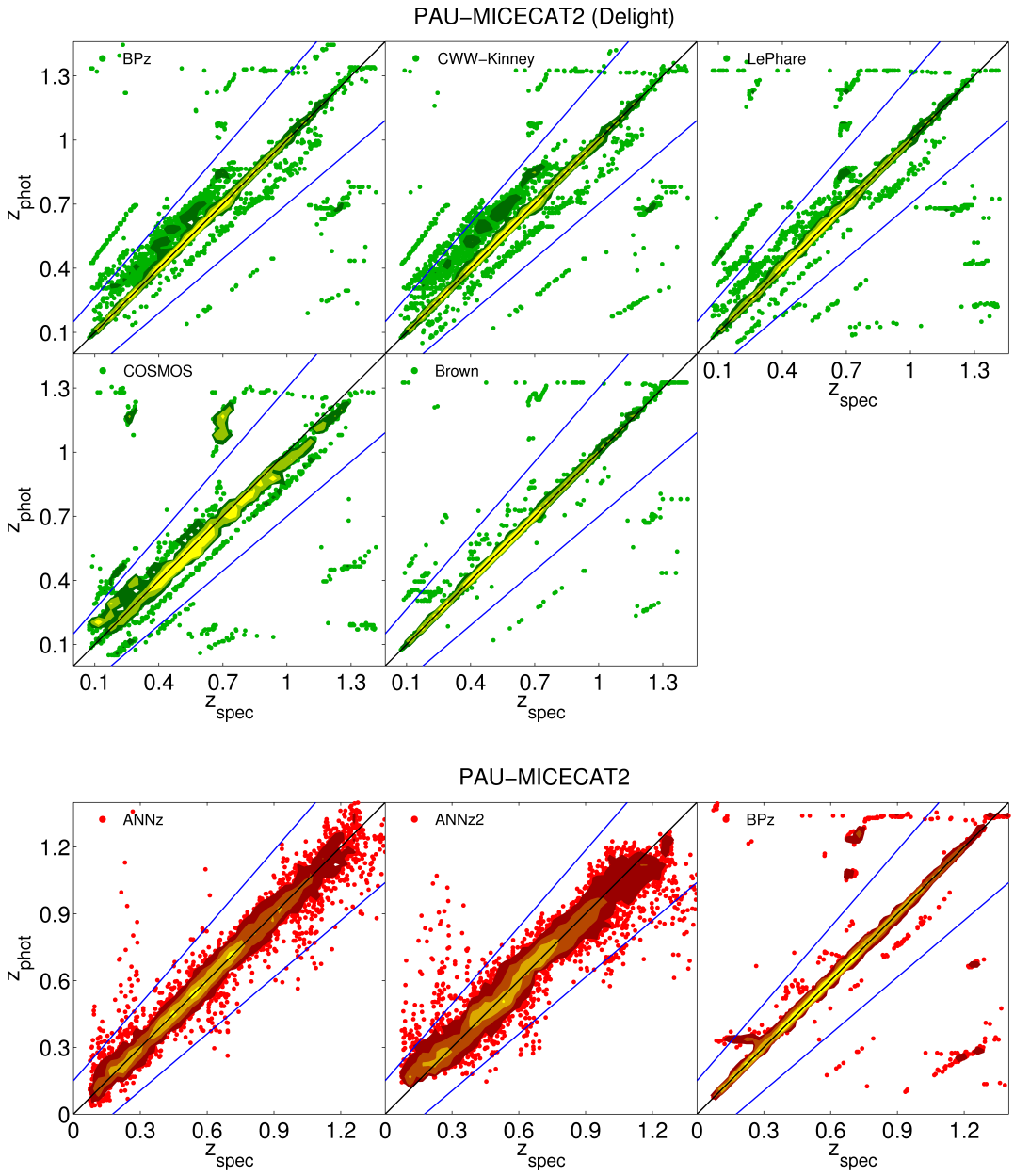

Figure 79: Photo-z v.s. spec-z performance of DELIGHT (top), ANNZ, ANNZ2 and BPZ (bottom), respectively, when trained with 40 bands on the PAU-MICE2 Simulated Sample. The 5 panels on the top show the performances of DELIGHT when using the BPZ, CWW-Kinney, LEPHARE, COSMOS and Brown templates, respectively. Note that the BPZ run here uses the Brown templates.

Table 30: Photo-z performances of DELIGHT using different templates for the PAU-MICE2 Simulated Sample, with performance metrics σ_{RMS}, σ_{68}, and η_{out}. Also shown are the performances of ANNZ, ANNZ2 and BPZ for comparison.

Code	**Template**	σ_{RMS}	σ_{68}	η_{out} (%)
DELIGHT	BPZ	0.0646	0.0064	3.50
DELIGHT	CWW-Kinney	0.0695	0.0246	4.69
DELIGHT	LEPHARE	0.0999	0.0058	6.78
DELIGHT	COSMOS	0.1134	0.0429	9.33
DELIGHT	Brown	0.0533	0.0024	1.99
ANNZ	-	0.0367	0.0194	0.89
ANNZ2	-	0.0525	0.0331	2.01
BPZ	Brown	0.0725	0.0027	3.38

Im), we can safely say that the reduction of this scatter is credited to the interpolation. However from Table 30 LEPHARE actually has a larger scatter due to the higher density of galaxies at certain off-diagonal areas on the figure, this is due to LEPHARE templates only having one starburst template, compared to 4 and 6 in the BPZ and CWW-Kinney templates.

The result using COSMOS templates is surprisingly worse than expected: we expected it to do better than the rest since the catalogue was constructed using COSMOS templates. We believe that this is due to the two extra modifications on the COSMOS templates when generating PAU-MICE2: firstly, each adjacent template were linearly interpolated, further creating 10 templates between these templates; and secondly, extinction laws from Prevot et al. (1984) and Calzetti et al. (1994) were additionally applied to spiral galaxies and the starburst templates. We initially expected that the GP would have been able to automatically correct for extinction and mimic template interpolation with the help of the training set, however our results show that the dependence on good and representative templates is still crucial for DELIGHT to obtain good photo-z's in the context of a narrowband survey.

Despite the metrics showing that the results of COSMOS templates is the worst of them all with an exceptionally higher σ_{68}, we however see certain streaks which were present in all 3 previous templates missing here as shown in Figure 79. Since the elliptical and spiral templates of COSMOS are obtained from Polletta et al. (2007) instead of CWW, and the fact that COSMOS uses 12 templates from Bruzual and Charlot (2003), we suspect this to be the cause of the disappearance of these streaks.

The results of the Brown et al. (2014) templates, however, are exceptionally good when compared to the rest. This may be due to the fact that Brown templates are generated from real galaxies which better represent errors of galaxies, and also due to the large number of templates it has, being a total of 129 templates, more than half of the number LEPHARE templates, which is in second place. With the Brown templates the we have $\sigma_{68} \sim 0.0024(1+z)$, lower than the expected requirement even without any quality cuts on the sample. We see that the results of BPZ running with the same templates achieve similar performance too from Table 30.

Therefore in this section we conclude that Brown templates are the best choice for DELIGHT to use to obtain photo-z's for the PAU narrowband survey. We proceed to further optimise DELIGHT in the following section.

8.2.3 *Application: Estimating the Luminosity Function*

In this section we attempt to test the quality of the photo-z's produced by ANNz, ANNz2 and DELIGHT by estimating the luminosity function. The luminosity function as discussed in Section 2.1.1 can be seen as the number density of galaxies greater than a certain absolute magnitude. This is not as simple as creating a cumulative distribution of absolute magnitudes, since in an apparent magnitude (flux) limited survey, we expect more luminous galaxies to be detected at larger distances, therefore this bias has to be taken into account.

In Equation 38, we have shown that the differential form of the luminosity function was parametrized by Schechter (1976) as $\Phi(M)\,dM = \phi_\star x^\alpha e^{-x}\,dx$, where $x = 10^{-0.4(M-M_\star)}$, M is the absolute magnitude, and $\{\phi_\star, \alpha, M_\star\}$ to be estimated by data fitting. Integrating this equation would essentially give us the number density of galaxies above a certain absolute magnitude. For a quick analysis, we however chose a simpler way to estimate the luminosity function, by using the so-called *V_{max} method* (Schmidt 1968; Felten 1976).

$V_{\max}$ is defined to be the maximum comoving survey volume where a galaxy with such a luminosity would have been detected and included in the flux limited catalogue, represented by

$$V_{\max} = \frac{\Omega}{3}\left[d_C(z_{\max})^3 - d_C(z_{\min})^3\right], \tag{117}$$

where Ω is the solid angle, and d_C the comoving distance (Equation 23) at the maximum or minimum redshift ($z_{\max}$ and $z_{\min}$) the galaxy with said absolute magnitude is to be included in the catalogue. If we assume that the catalogue does not have a bright limit ($d_C(z_{\min}) = 0$), we can write down $d_C(z_{\max})$ and $d_C(z)$ in relation to M_b in band b by

$$M_b = m_{\text{faint},b} - 5\log_{10}\left[(1+z_{\max})d_C(z_{\max})\right] - 25 - K_b(z_{\max}), \tag{118}$$

$$M_b = m_b - 5\log_{10}\left[(1+z)d_C(z)\right] - 25 - K_b(z), \tag{119}$$

where m_b is the apparent magnitude and K_b the K-correction. For simplicity, we choose to ignore the K-correction terms (or assert that $K_b(z_{\max}) \approx K_b(z)$), we can then combine both equations to get

$$(1+z_{\max})d_C(z_{\max}) = 10^{\frac{m_{\text{faint}} - m + 5\log_{10}((1+z)d_C(z))}{5}}. \tag{120}$$

Once we find $z_{\max}$ numerically, we can estimate $d_C(z_{\max})$ to find $V_{\max}$, normalise it by $\sum V_{\max}$, it can then form a simple and unbiased estimator of the galaxy luminosity function,

$$\Phi(M)\,\Delta M = \sum \frac{1}{V_{\max}}. \tag{121}$$

This gives us a number distribution of M in bins of ΔM, but each galaxy weighted by the inverse of the total volume it could occupy when being observed. In simple words, if we have the absolute magnitude, apparent magnitude, redshift and comoving distance of each galaxy in a catalogue in a certain band filter b, we are able to estimate the luminosity function $\Phi(M_b)$ in small bins of ΔM_b.

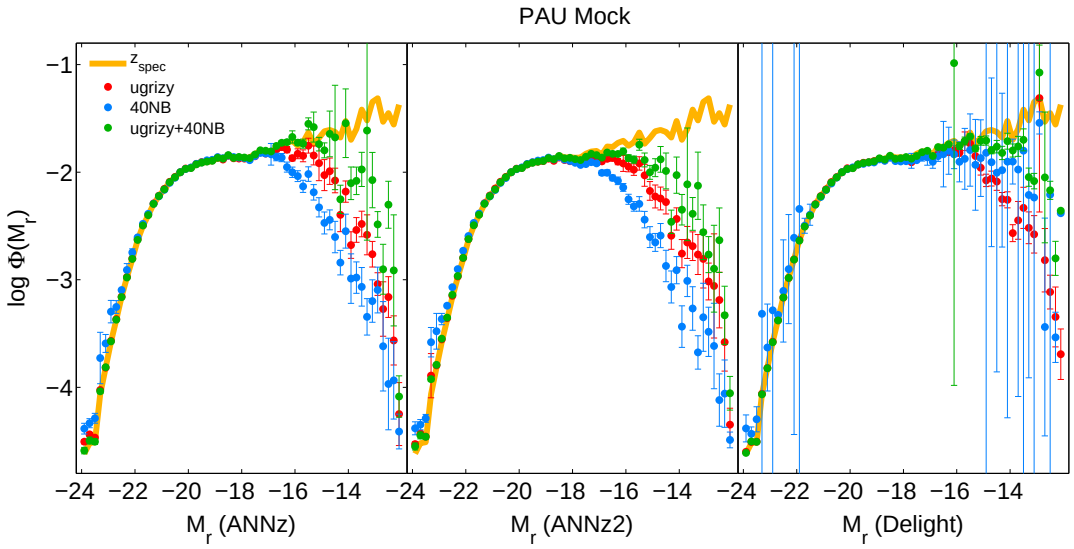

Figure 80: The luminosity function estimated using the $V_{\max}$ method for the PAU Mock Sample, comparing the results by ANNZ (left), ANNZ2 (middle) and DELIGHT (right) when trained using 6 broadbands, 40 narrowbands, and both, respectively.

In this section we utilise the spec-z's and photo-z's produced by ANNZ, ANNZ2 and DELIGHT from the PAU Mock Sample to calculate $V_{\max}$ for each object and compare the luminosity functions produced by each code and number of filters. We select a bin of $\Delta M_r = 0.2$ in the r-band, and placing a faint apparent magnitude limit of $m_r < 22.6$. The different luminosity function plots using different photo-z codes are shown in Figure 80, where the error bars are calculated via propagation of errors.

In comparison with the luminosity plotted using $z_{\rm spec}$, we find that all photo-z's accurately depict the luminosity function up till $M_r < -17$, and the luminosity function is underestimated beyond this point, it is most extreme when using only the 40 narrowbands. When all 46 bands are used, the faint end of the luminosity function is recovered further up to $M_r < -16.5$. The results for the two machine learning codes ANNZ and ANNZ2 are very similar, we see that the fits with only 40 narrowbands produced the worst results, while with the help of broadbands the faint end of the luminosity function is less underestimated. A small bias is also seen at the bright end around $M_r \sim -23$. The 40-narrowband fit for DELIGHT has very large error bars due to the large error outputs of its photo-z produced, despite that it produces better results than the machine learning codes. Overall the fit using 46 bands with ANNZ and DELIGHT are comparable, with DELIGHT having more points clustered closer to the $z_{\rm spec}$ line, arguably recovering the luminosity function close to $M_r < -14$. Quantitatively, the mean biases of the luminosity function fits for the codes ANNZ, ANNZ2 and DELIGHT (using 46 bands) are found to be -0.1688, -0.2585 and -0.0720 respectively (in units of $\log \Phi(M_r)$), where the bias of DELIGHT is shown to be less than half of the bias of ANNZ.

Overall we conclude that in the context of PAU, the photo-z's produced by DELIGHT using 46 bands is found to be the best choice for luminosity function estimation up to around $M_r < -16$. However, we expect that the underestimation at the faint end of the luminosity function would be less severe if (1) K-correction was accounted, since this correction is significant and non-

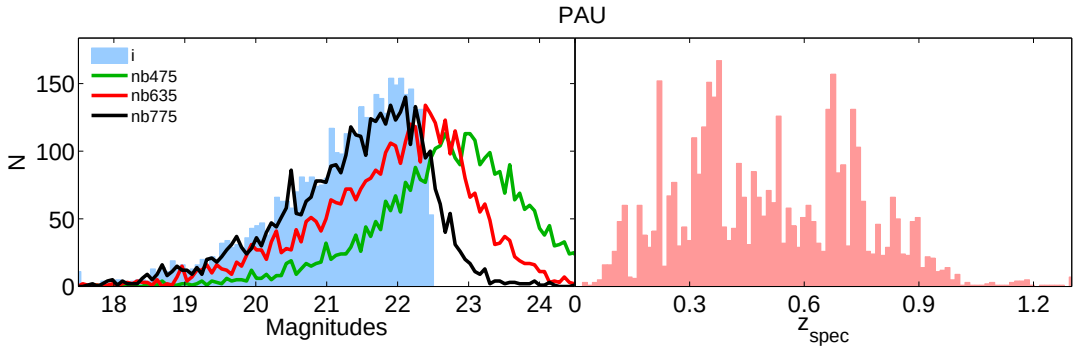

Figure 81: Distribution of i, $nb475$, $nb635$, $nb775$ (left) and z_{spec} (right) for the PAU (Real) Sample.

negligible especially for higher redshift galaxies; and if (2) the luminosity function is estimated separately by galaxy type, since different galaxy types have different absolute magnitude limits (Ilbert et al. 2004).

We note that while there are many different methods to estimate the luminosity function (see Willmer 1997), the V_{max} method remain one of the easiest to implement, it simultaneously produces the shape and normalisation for the luminosity function, and it produces relatively good results for any survey with a smooth $n(z)$ (Willmer 1997). The V_{max} method suffers from the fact that it can be biased by the presence of large scale structures, however we assert that its use for the PAU Mock Sample is appropriate and sufficient.

8.3 PHOTOMETRIC REDSHIFTS ON THE PAU EARLY DATA RELEASE

Having tested and optimised DELIGHT on simulated galaxy samples, we now progress to use real data to assess its general performance, both on the point estimate and PDF. Here we construct the PAU (Real) Sample, cross-matched the forced-aperture coadded photometry from PAU (40 narrowbands) with the 3″ fixed-aperture photometry ($uBVriz$ broadbands) from COSMOS (Laigle et al. 2016) and spectroscopic redshifts from zCOSMOS (Lilly et al. 2007) to form our training and testing sample. The reader could refer to Section 4.1.5, Section 4.1.6 and Section 4.3.7 for full descriptions on PAU, COSMOS and zCOSMOS surveys.

The sample selection is as follows. For spectroscopic redshift, we select redshift qualities with confident class `3<CC<5`; for COSMOS we select objects which are galaxies (`TYPE=0`) and with the Kron i-band magnitude $i_{\text{auto}} < 22.5$; and for PAU since the photometry is currently still constantly being updated, we chose to use photometry from `PRODUCTION_ID=697` as it would allow us to compare with the early photo-z results of BCNZ2 (Eriksen et al., in prep.). We also ensure that the galaxies are detected in all of the 40 narrowbands (`N_BANDS=40`). Cross-matching these samples via sky position yields a sample of 8100 objects, divided into half for training and the other half for testing. The distribution of i and z_{spec} of this sample is shown in Figure 81, we note that the incompleteness in i is due to the selection effects when cross-matched to zCOSMOS spectroscopic redshifts.

8.3.1 *Calibrating Fluxes Between Broad and Narrow Bands*

Just as the analyses in Section 8.2.1, we intend to produce photo-z's for this sample using 40 narrowbands and 6 broadbands. As PAU currently do not have broadband fluxes, we match each PAU galaxy to the COSMOS field to obtain *uBVriz* broadbands. However, the fluxes from the two surveys were measured differently: the COSMOS *uBVriz* fluxes were measured using a fixed 3″ aperture flux, while the 40 PAU fluxes were measured using an aperture which covers 62.5% of the object flux[1]. Therefore, the fluxes from COSMOS need to be calibrated in order to match the fluxes of the narrowbands to be trained together.

The calibration process is done in two steps: first a calibration per galaxy, then a calibration per band. In the first step, we calibrate the fluxes in the training set by finding the best fit Brown template to each galaxy using only 40 narrowband fluxes from PAU. With the best fit template for each galaxy and their spectroscopic redshifts in hand, we use DELIGHT to predict the broadband fluxes according to their bands *uBVriz* (note that this technically cannot be done on a target set without spectroscopic redshift, we will discuss alternative methods to this in Section 8.3.2). Here we assume the predicted flux to be the true flux of the broadbands and that the Brown templates have represented most if not all galaxy spectral types. This predicted flux $F_{\mathrm{p},b}$ for each band b is then compared to the original COSMOS flux $F_{0,b}$, as shown in the top row of Figure 82 for a sample of u, V and z. We can see that, while the correlation is high for each band, there are offsets that need to be corrected.

Next, we take the ratio between the predicted flux and the original flux, $R_{\mathrm{g},b} = \frac{F_{\mathrm{p},b}}{F_{0,b}}$ for each band b. Assuming that this flux ratio should be almost the same across each band for each galaxy, we take the weighted mean of $R_{\mathrm{g},b}$ as follows:

$$\bar{R}_{\mathrm{g}} = \frac{\sum_b^{n_b} \frac{1}{\sigma^2_{R_{\mathrm{g}},b}} R_{\mathrm{g},b}}{\sum_b^{n_b} \frac{1}{\sigma^2_{R_{\mathrm{g}},b}}}, \tag{122}$$

where n_b is the number of bands and $\sigma^2_{R_{\mathrm{g}},b}$ the variance of $R_{\mathrm{g},b}$, which can be derived from the original and predicted flux variances $\sigma^2_{F_0}$ and $\sigma^2_{F_{\mathrm{p}}}$ by propagation of error. Therefore we now have a factor $\bar{R}_{\mathrm{g}}$ (which we shall call 'the broadband flux offset factor per galaxy', or just 'offset factor' for short) for each galaxy which we could use to obtain a calibrated flux per galaxy for each band, $F_{\mathrm{g},b} = \bar{R}_{\mathrm{g}} F_{0,b}$. This calibration is motivated by the fact that each galaxy having different sizes has different fluxes measured. In fact we find a positive correlation between $\bar{R}_g$ and r_{50}, the semi-major axis of the galaxy where 62.5% flux is measured, gaining an equation of $\bar{R}_{\mathrm{g}} = 0.0145 r_{50} + 0.4112$, and a correlation coefficient of $r_{\mathrm{corr}} = 0.878$. This is visualised in Figure 83, with a distribution of $\bar{R}_{\mathrm{g}}$, having a median value of 0.6931, meaning that on average COSMOS measures more flux for each galaxy than PAU.

We then proceed to the next step where we calibrate the broadband magnitude offsets for each band by doing a weighted least-squares fit between $F_{\mathrm{p},b}$ and $F_{\mathrm{g},b}$ for each broadband. A weighted fit is required here since we expect that objects that are brighter to have relatively lower variances, and by accounting for the variances of objects the fainter objects would be

1 This is true for the forced-aperture photometry from `PRODUCTION_ID=697`. At the time of writing this has increased to 75%, and the value may still change in the future.

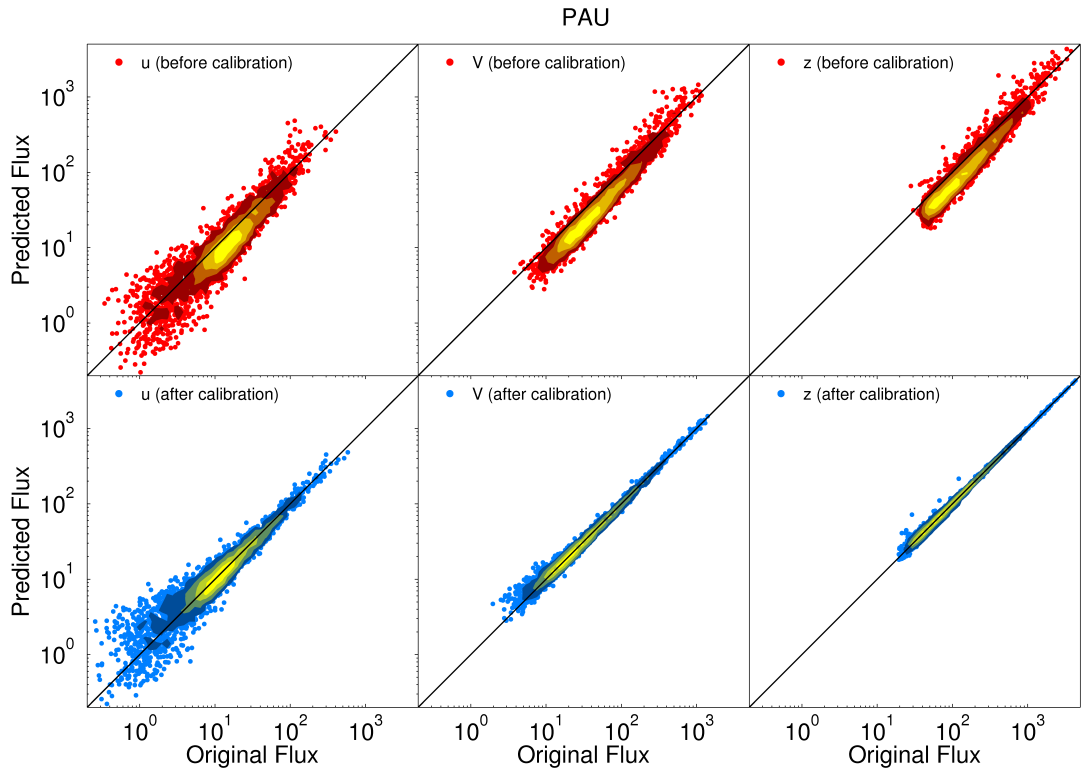

Figure 82: Comparison between the broadband fluxes predicted by DELIGHT and the original COSMOS $3''$ fluxes before calibration (top row) and after calibration (bottom row), shown for the u-band (left), V-band (middle) and the z-band (right) for objects in the PAU trainings set. The fluxes are measured in units of 100 nJy.

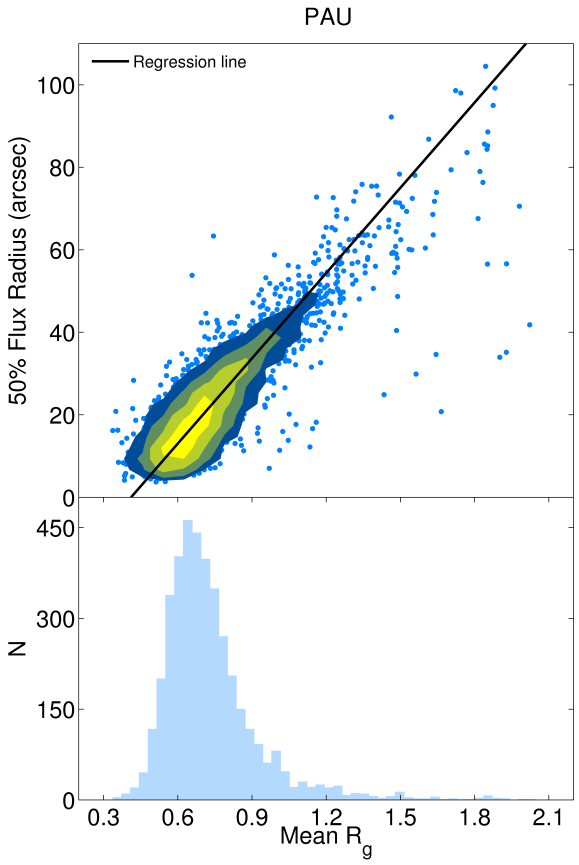

Figure 83: $\bar{R}_g$ v.s. the 50% flux radius r_{50} (top) for each galaxy in the PAU training set, and the tabulated distribution of $\bar{R}_g$ (bottom).

Table 31: Best fit constants m_b and c_b for each band b through a weighted least-squares fit for the equation $\ln(F_{\mathrm{p},b}) = a_b \ln(F_{\mathrm{g},b}) + c_b$.

b	a_b	c_b
u	1.0048 ± 0.0002	0.0641 ± 0.0008
B	0.9998 ± 0.0003	0.239 ± 0.001
V	1.0137 ± 0.0002	-0.091 ± 0.001
r	1.0222 ± 0.0002	-0.0395 ± 0.0008
i	1.0111 ± 0.0001	0.0265 ± 0.0006
z	1.00090 ± 0.00005	-0.0169 ± 0.0003

upweighted. We chose a simple linear equation of $\ln(F_{\mathrm{p},b}) = a_b \ln(F_{\mathrm{g},b}) + c_b$, where a_b and c_b are constants to be optimised for each band. These constants are calculated through the weighted least-square fit and are tabulated in Table 31.

As expected from the table, the values of m_b and c_b are very close to 1 and 0 respectively since the calibrated flux per galaxy $F_{\mathrm{g},b}$ is already very close to the predicted flux $F_{\mathrm{p},b}$. The final calibrated flux for each broadband is therefore $F_{\mathrm{gb},b} = e^{(a_b \ln F_{\mathrm{g},b} + c_b)}$. The second row of Figure 82 shows the correlation between the final calibrated and predicted fluxes for u, V and z, and the improvement between the top and bottom panels are visible.

The hyperparameters of DELIGHT were further tweaked to allow the production of more precise results: the locality of the model predictions in redshift (`zPriorSigma`) and luminosity (`ellPriorSigma`) were decreased from 0.2 to 0.1 and 0.5 to 0.2 respectively; a 3% and 6% flux were added in quadrature to the flux variances of the narrowbands and broadbands respectively to account for other flux errors both from the data and the model (values estimated via trial and error).

ANNZ, ANNZ2 and BPZ were optimised to produce results to be compared with DELIGHT too. However due to lack of spectroscopic data, ANNZ and ANNZ2 suffer from only having about 2000 objects each for training and validating, which is the minimum threshold for the number of training objects for decent broadband photo-z results (Collister and Lahav 2004; Bonfield et al. 2010). We expect that the threshold for a sample with large number of narrowbands should be even higher, therefore their results may be affected. BPZ was optimised in a similar manner, tested using the same calibrated flux , Brown templates and COSMOS prior as mentioned in Section 8.2.2. Photo-z PDFs for ANNZ2, BPZ and DELIGHT were produced with a binning size of $\delta z = 0.005$.

8.3.2 *Results*

The final results for all 4 codes are shown in Table 32 and visualised in Figure 84. If trained with only 6 broadbands, we see ANNZ performing the best as expected when compared to all other codes. Here we also see ANNZ performing worse than ANNZ2 at 40 and 46 bands, seemingly suggesting that for real data ANNZ2 is able to cope with a large number of inputs better than ANNZ.

The real spotlight however is on DELIGHT: we see it performing extremely well at 46 bands after calibration, reaching a 68th percentile error of $\sigma_{68} = 0.0082(1+z)$, slightly better than

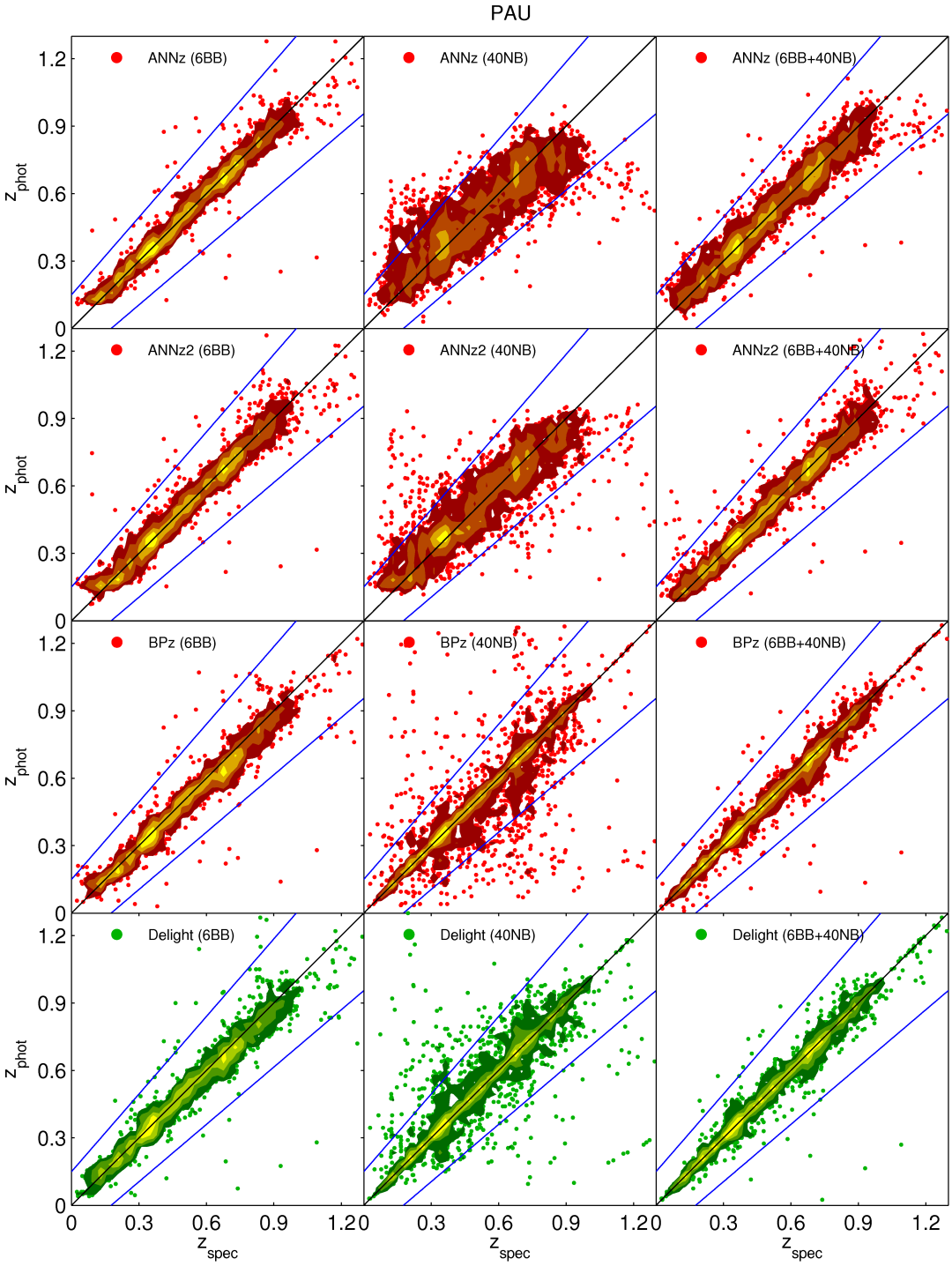

Figure 84: Photo-z v.s. spec-z performance of ANNz (top row), ANNz2 (second row), BPz (third row) and DELIGHT (last row) when trained with 6 bands (left column), 40 bands (middle column) and 46 bands (right column), respectively, on the PAU Sample.

Table 32: Photo-z performances of ANNZ, ANNZ2, BPZ and DELIGHT on the PAU Sample, shown with the respective number of training bands n_b and performance metrics $\sigma_{\rm RMS}$, σ_{68} and $\eta_{\rm out}$. Also shown in the table are the metrics for DELIGHT before calibration, and the photo-z results of BCNZ2 (Eriksen et al., in prep.).

Code	n_b	$\sigma_{\rm RMS}$	σ_{68}	$\eta_{\rm out}$ (%)
ANNZ	6	0.0298	0.0183	0.47
ANNZ2	6	0.0362	0.0243	0.64
BPZ	6	0.0357	0.0213	0.89
DELIGHT (no cal.)	6	0.0502	0.0441	0.99
DELIGHT	6	0.0343	0.0226	0.81
ANNZ	40	0.0856	0.0705	7.63
ANNZ2	40	0.0803	0.0551	5.09
BPZ	40	0.1111	0.0244	6.96
DELIGHT	40	0.0723	0.0150	4.49
ANNZ	46	0.0433	0.0323	1.04
ANNZ2	46	0.0366	0.0225	0.77
BPZ	46	0.0289	0.0085	0.59
DELIGHT (no cal.)	46	0.1341	0.0447	6.64
DELIGHT	46	0.0274	0.0082	0.49
BCNZV2	46	0.0332	0.0089	0.81

BPZ (especially at the very low redshift region), and also significantly better than the run with 40 bands. This proves that the flux calibration is successful. The improvement due to the flux calibration is highlighted in Figure 85, where a lot of scatter has been fixed. An interesting observation here is that the results for 6 bands improved with the calibration as well, which suggests that fluxes with matched-apertures produce better photo-z's than fluxes with fixed aperture sizes when using template methods.

Although we see great results for DELIGHT after calibration for the training with 46, we note that the calibration by galaxy for the testing set is made possible due to it having spec-z's in the first place. We attempt to simulate the realistic situation when we do not have spec-z's in the testing set with the following three methods: (1) we skip the calibration per galaxy for both the training and testing, instead only calibrating the offsets between the broadband fluxes, and use this as the final calibrated broadband flux $F_{{\rm b},b}$; (2) we predict $F_{{\rm p},b}$ using the photo-z produced from the training with 40 bands instead of the spec-z of the testing set, then use it to calculate $\bar{R}_{\rm g}$ and $F_{{\rm gb},b}$; (3) we use the relationship between $\bar{R}_{\rm g}$ and r_{50} in the training set (Figure 83) to estimate $\bar{R}_{\rm g}$ for each galaxy in the testing set, and use this to calculate $F_{{\rm gb},b}$. For objects without the value of r_{50} in the catalogue, we set $\bar{R}_{\rm g} = 0.6931$ which is the median value of $\bar{R}_{\rm g}$. The newly calibrated broadband fluxes from these three methods are trained together with the 40 narrowbands, and the performance metrics for these methods are summarised in Table 33.

From the table, the results of method 1 highlights the importance of flux calibration per galaxy, seeing that without such calibration the σ_{68} almost doubles in value. The results for method 2 and 3 are interesting because they are both quite close to our default calibration mentioned above. The results of method 2 show that spec-z can be replaced by the 40 band photo-z to predict broadband fluxes for calibration without having severe degradation in performance. We have

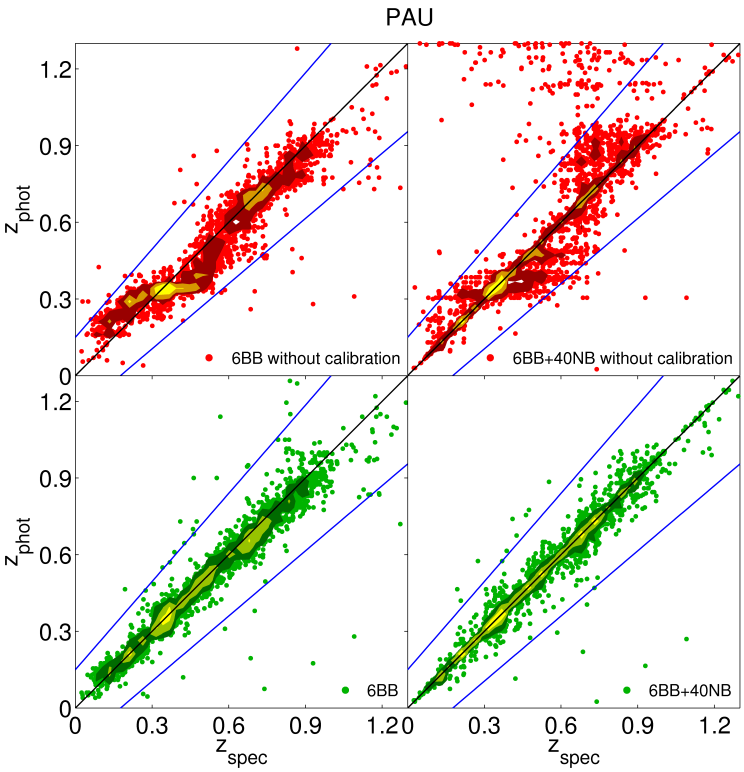

Figure 85: Photo-z v.s. spec-z performance for the PAU Sample, comparing the results of DELIGHT for 6 bands (left column) and 46 bands (right column) respectively, with the top and bottom row showing the results before and after calibration.

Table 33: Photo-z performance of DELIGHT when trained using 46 bands, comparing the 3 different realistic situations as mentioned in the third paragraph of Section 8.3.2. This table summarises the redshift used to predict the broadband fluxes $F_{\mathrm{p},b}$, the method to estimate the offset factor $\bar{R}_{\mathrm{g}}$, and whether the flux offset between the broadbands are calibrated. The last three columns show their respective performance metrics σ_{RMS}, σ_{68} and η_{out}.

Description	**Redshift to predict $F_{\mathrm{p},b}$**	**Estimating $\bar{R}_{\mathrm{g}}$**	**Offset between bands**	σ_{RMS}	σ_{68}	η_{out} (%)
No calibration	-	-	-	0.1341	0.0447	6.64
Method 1	-	-	calibrated	0.0374	0.0173	0.89
Method 2	z_{phot} (from 40 bands)	$\bar{R}_{\mathrm{g}} = F_{\mathrm{p},b}/F_{\mathrm{o},b}$	calibrated	0.0307	0.0096	0.62
Method 3	-	via correlation with r_{50}	calibrated	0.0291	0.0090	0.47
Default	z_{spec}	$\bar{R}_{\mathrm{g}} = F_{\mathrm{p},b}/F_{\mathrm{o},b}$	calibrated	0.0274	0.0082	0.49

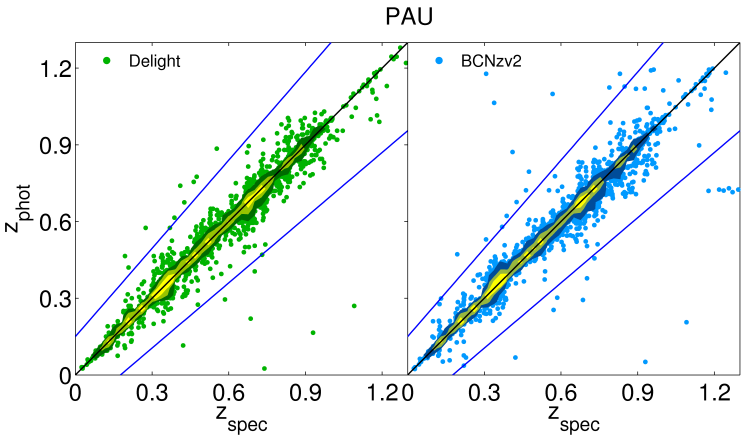

Figure 86: Photo-z v.s. spec-z performance for the PAU Sample, comparing the results of DELIGHT (left) and BCNZV2 (right) respectively, both trained on 46 bands.

also checked that the predicted fluxes using the two different redshifts are highly correlated, meaning that the best fit template selected for each galaxy to predict their fluxes are similar despite the use of two different types of redshift. However the result for method 3 is surprising since we mentioned in Section 8.3.1 that we only have a fairly strong correlation between $\bar{R}_g$ and r_{50} from the training set (correlation coefficient $r_{\rm corr} = 0.873$), and yet this relationship produced photo-z with a performance surpassing method 2. This is good news since the half-light radius of galaxies usually comes for free when fluxes are estimated, this morphological parameter can be well utilised as a fast, simple yet accurate method to calibrate the broadband fluxes.

Next, we are also interested to compare the photo-z performances between DELIGHT and BCNZV2 (Eriksen et al., in prep.). BCNZV2 is a template-based photo-z code purpose-built for the PAU Survey photo-z analysis. The analysis incorporates a relative calibration between the broad and narrowband fluxes, as well as a scaling on a galaxy to galaxy basis. As for templates, BCNZV2 uses the COSMOS templates but additionally incorporated emission lines using a similar method to Ilbert et al. (2009) as well as extinction laws considered for some templates. We recommend the reader to read the upcoming paper for further details. Here we cross-matched our testing sample with photo-z with their first photo-z data release (internal release), and find that our results are comparable with theirs, in fact having slightly less scatter, as seen in Figure 86.

We also compare the σ_{68} of both DELIGHT and BCNZV2 while systematically cutting the samples based on their odds parameters, keeping only objects with the best 90%, 80%, 70%, 60%, 50% and 40% odds values, and plot them with respect to the $i_{\rm auto}$ cumulatively, as shown in Figure 87. Here we see that for our default flux calibration procedure, while BCNZV2 goes below the $\sigma_{68} = 0.0035(1+z)$ line at 40% odds cut, we see DELIGHT still slightly above the line. We also see DELIGHT performing better than BCNZV2 at the brighter end. At the fainter end, DELIGHT seems to perform better overall, however the odds produced by DELIGHT seem not able to remove bad photo-z's as efficiently as BCNZV2. The figure also shows that our results using method 3 are very similar to our default calibration at lower percentage odds cuts. In summary, DELIGHT performs better for the metrics at a larger percentage of the sample, however BCNZV2 performs better with odds cuts further down the sample size.

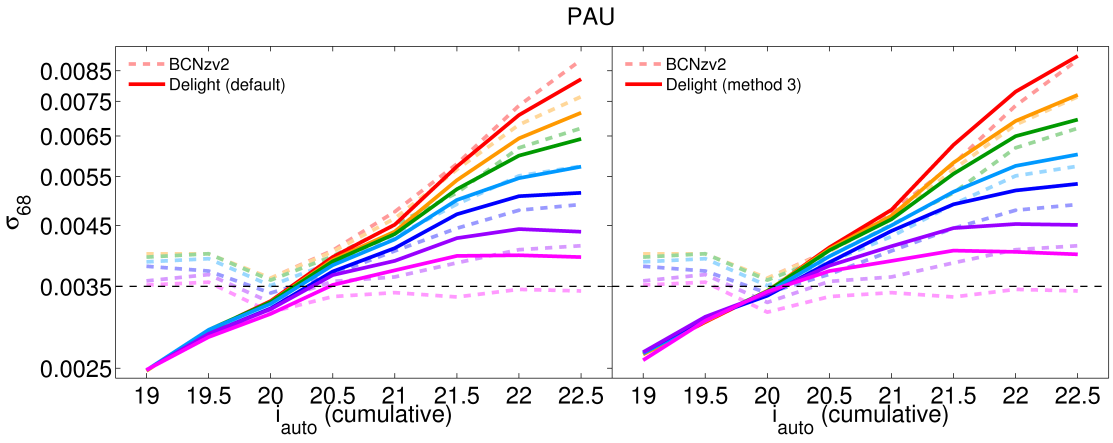

Figure 87: σ_{68} for galaxies brighter than i_{auto} for the PAU testing set, comparing the results from DELIGHT trained with 46 bands (solid lines) and BCNZV2 (dotted lines). The colours denote quality cuts on the sample based on the odds parameter, with red being the full sample, and the rest keeping objects with the best 90% (orange), 80% (green), 70% (light blue), 60% (dark blue), 50% (purple) and 40% (magenta) odds values, respectively. The black line shows $\sigma_{68} = 0.0035(1+z)$, which is the nominal photo-z requirement as expected by the PAU Survey. The plot on the left shows the results of our default flux calibration, while the one on the right uses the flux calibration using method 3 mentioned in Table 33.

We note here however that the results of BCNZV2 shown in this work are not final: a newer data set and with the implementation of new scaling and calibration methods has been released internally recently, and we further expect that the public release of BCNZV2 photo-z's to have even better results. Still, we find the results by DELIGHT promising and anticipate further room for improvement.

Finally, we also assess the quality of the photo-z PDFs produced for this sample, especially in the context of a narrowband survey. Figure 88 shows the QQ plot for ANNZ2, BPZ and DELIGHT, respectively, and overall we see that ANNZ2 and DELIGHT have lines close to the diagonal, showing that overall they have produced photo-z PDFs which represent the set well. Noticeably BPZ produces PDFs that are too broad in this case, giving the large Z-shaped curve. Taking a closer look, the photo-z PDFs by ANNZ2 are slightly broader than expected, while DELIGHT slightly narrower than expected, with more objects having their photo-z's over-predicted than under-predicted.

Overall, we conclude that DELIGHT not only produces good photo-z point estimates but also good PDFs. Note that in this analysis we do not discuss the quality of $n(z)$ produced by the different codes for several reasons. Firstly, the number of objects in the testing set is small, resulting in the $n(z)$ distribution being dominated by noise. Secondly, as the testing set is situated in the relatively small COSMOS field, the actual spec-z distribution is dominated by local large-scale structure which are not representative in the large scale, making the evaluation of $n(z)$ difficult. Lastly, the n_{KS} values produced by the PDFs are exceptionally high due to the small redshift binning size chosen.

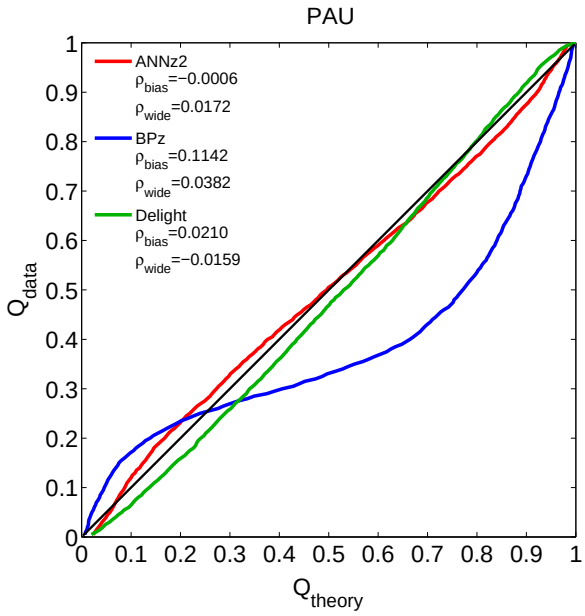

Figure 88: *QQ* plot for the photo-z PDFs produced for our PAU testing set, comparing the performances between ANNz2 (red), BPz (blue) and DELIGHT (green).

8.3.3 *Analysis on Missing or Undetected Bands*

As mentioned earlier, with the large number of bands being observed in the PAU Survey, we expect the number of missing or undetected bands to be higher than in the usual broadband surveys. Besides, observations at the edges of the survey field are bound to have certain number of bands undetected due to the PAU Survey observation strategy, and we attempt to see how much these undetected bands would affect the photo-z performance.

Using the same training and testing set and settings as above, we trained DELIGHT using only 40 narrow bands, and we then evaluate this on the testing set not with all 40 bands, but instead using 38, 36, 35, 32, 30, 27, 20, 13 and 10 bands, respectively. In each case, the discarded bands are not removed randomly but alternating a fixed number of bands (e.g. removing 1 in every 4 bands, 1 in every 8 and etc). Photo-z's are obtained for these runs and the their respective performance metrics are shown in Table 34, and the values of $\sigma_{\rm RMS}$ and σ_{68} across bands are shown in Figure 89.

When compared to our previous results in Table 32, we see that the removal of 1 in 5 bands (i.e. 32 bands) in the testing set has its σ_{68} reaching a value which is equivalent to that of when trained and tested on 6 broadbands. The σ_{68} doubles when it reaches 27 bands, and it essentially reaches a photo-z quality equivalent to our uncalibrated broadband results when the bands are halved. When compared to machine learning codes for training with 40 bands, the σ_{68} of DELIGHT is still way above the results of ANNz and ANNz2 even when the number of tested bands is halved, however its scatter and outlier fraction are comparable to both. We also notice from Figure 89 that the reduction of bands used in training affected the σ_{68} more than the $\sigma_{\rm RMS}$, showing the main impact of multiple narrowbands on the σ_{68} of photo-z's rather than the scatter.

Table 34: Photo-z performance of DELIGHT for the PAU Sample, when trained with 40 bands but tested on fewer number of bands n_b, shown with their respective performance metrics $\sigma_{\rm RMS}$, σ_{68} and $\eta_{\rm out}$. Highlighted in red shows the threshold where the σ_{68} drops below the quality of photo-z's produced by training with only the 6 $uBVRiz$ broadbands.

n_b	$\sigma_{\rm RMS}$	σ_{68}	$\eta_{\rm out}$ (%)
40	0.0723	0.0150	4.49
38	0.0720	0.0185	4.83
36	0.0753	0.0215	4.97
35	0.0769	0.0212	5.31
32	0.0793	0.0246	5.62
30	0.0806	0.0275	6.54
27	0.0781	0.0310	6.03
20	0.0883	0.0392	7.86
13	0.0925	0.0542	9.42
10	0.1027	0.0667	11.08

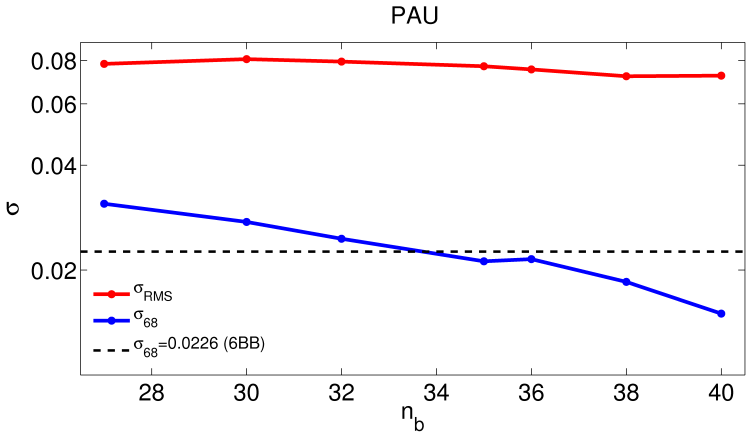

Figure 89: Performance metrics $\sigma_{\rm RMS}$ (red) and σ_{68} (blue) across the number of testing bands n_b when trained with 40 bands using DELIGHT on the PAU Sample. The black dashed line shows the σ_{68} for the training with 6 $uBVriz$ broadbands only.

Therefore we recommend that any photo-z derived using less than 32 bands are deemed not up to the expected performance of PAU. This means that observations near the edge of the field with less than 32 bands observed could be safely ignored as they are practically not going to produce satisfactory photo-z's. However this assumption only applies to the situation when no broadband information is available: we believe that the threshold might get lower when broadbands are included in this analysis. Besides, we could potentially use the missing band prediction method mentioned earlier to see if the results improves. The study on the impact of missing or unobserved bands in the situation when broadbands are available shall be left as a potential future work to be ventured and expanded.

8.4 FUTURE PROSPECTS OF DELIGHT ON NARROWBAND SURVEYS

Thus far we have seen that DELIGHT performed really well in both simulated and real galaxy samples. We see that with good templates, well optimised hyperparameters and well calibrated broadband and narrowband fluxes, DELIGHT is able to produce high quality photo-z's. With such high quality photo-z's produced, we find that catastrophic outliers are now easier to be identified (e.g. the same objects in the lower right corner of each plot in Figure 84) and could potentially be individually investigated to see if a different template is needed or that the spectroscopic redshifts for these objects were catastrophic.

However, there is still a lot of room for improvement for DELIGHT. Firstly, DELIGHT can be tweaked such that emission lines could be incorporated on top of the continuum templates, as in LEPHARE and BCNZV2. To do so, however, may require a major overhaul of the code itself, but it is not an impossible task. Next, priors in template and redshift can also be included in DELIGHT to improve the results. In fact DELIGHT actually has such a function, we have tried to incorporate priors in our runs however we find results to be inconsistent each time, and the optimisation code could be looked into in the future.

Finally, the results of DELIGHT in this work (particularly the results on the real data) are mainly limited by the processing power and computing time within the working time frame. The redshift stepsize of $\Delta z = 0.005$ used in this work, although 2 times smaller than many current photo-z catalogues, is not precise enough to reach the requirement of $\sigma_{68} = 0.0035(1+z)$, the current stepsize should at least be 10 times smaller. However a finer stepsize for both the GP network and the redshift bin proved to be too much for the author's computer. The results may improve if more high-performance computing facilities are explored and utilised.

Part III

CONCLUSIONS

9

CONCLUSION AND OUTLOOK

For I am convinced that neither death nor life, neither angels nor demons, neither the present nor the future, nor any powers, neither height nor depth, nor anything else in all creation, will be able to separate us from the love of God that is in Christ Jesus our Lord.

Romans 8:38-39

9.1 CONCLUSION

In this thesis we have studied the effects of galaxy morphology on photo-z point estimates, PDFs and distributions of galaxies from various samples. We have explored using various different metrics to characterise the quality of photo-z PDFs and distributions. We have also compared the performances of 4 different codes (ANNZ, ANNZ2, BPZ and DELIGHT) on broadband and narrowband samples, optimised their settings, and even used them to produce photo-z catalogues for CS82, LSST and PAU. Here we summarise the 3 main research chapters in the following points:

1. Galaxy morphology is found to improve photo-z point estimates for all galaxy samples that we have used in this work, namely the MGS, colour selected samples (LRG, CMASS and LOWZ), Stripe-82 sample and the CS82 sample when reweighted with respect to the full CS82 Survey.

2. Surface brightness, angular radius and axis-ratio are the top 3 galaxy morphological parameters that improve photo-z point estimates the most. However, these galaxy morphological parameters should be used collectively when training to produce maximum improvement.

3. The degree of improvement in photo-z point estimates varies across different samples. Generally, improvement is found to be greater in spiral than in elliptical galaxy samples; greater in brighter than in fainter galaxy samples; greater when the number of bands in the survey is limited; and greater when the quality of photometry is low.

4. Galaxy morphology can be used in conjunction with quasar redshifts to improve photo-z's for a sample which has a weak star-galaxy separator. Galaxy morphology can be used with infrared magnitudes to further improve the photo-z's too.

5. Galaxy morphology is found to improve photo-z PDFs by increasing their average odds parameter, reducing their mean CRPS score, and reduce the width of the PDFs in the context of the reweighted CS82 sample. However the improvement is only marginal.

6. Galaxy morphology is found to not improve the overall photo-z distribution of a sample. However, it was found that with as few as 2 broadbands and morphology, an $n(z)$ distribution as good as using 5 broadbands can be produced. In the process, we find that the KS test is not a suitable metric for evaluating the performance of photo-z distributions.

7. Two state-of-the-art photo-z algorithms, ANNZ2 and DELIGHT have their performances compared on both a broadband survey (CS82) and a narrowband survey (PAU). Both codes have different advantages: ANNZ2 generally produces photo-z point estimates with a low scatter and outlier fraction, while DELIGHT with a lower σ_{68}. For a broadband survey with a large enough training set, ANNZ2 tends to produce better results, but for a narrowband survey DELIGHT does better with proper optimisation. Particularly in the context of the PAU Survey, using the largest and most realistic template set and calibration of fluxes between the photometry of PAU and COSMOS, competitive photo-z's can be produced.

9.2 FUTURE WORK AND OUTLOOK

In this work we have explored various ways to improve and characterise photo-z's, and in the process we find that there are a handful of experiments which could be extended and expanded as potential future work. In the following we list down a few areas where we think there is interesting future work to be pursued:

1. *What is 'improvement'?* Throughout this work we have continuously stated that galaxy morphology has improved photo-z's, but we have not defined what counts as an improvement, when is it considered significant, and how are different runs compared fairly. This question needs to be address from both the practical and statistical point of view, in order to really know how photo-z's are actually improved.

2. *The perfect photo-z PDF.* Although photo-z PDFs have been widely sought after, it is still not clear what the qualities of photo-z PDFs are required for key science cases. We may have attempted to characterise these PDFs in this work, but we have barely scraped the tip of the iceberg. Samples of galaxy PDFs can be simulated and studied, in order for photo-z algorithms to be optimised for best output. On the same topic, the characteristics of a perfect $n(z)$ could be explored too.

3. *Galaxy morphology with template-based codes.* DELIGHT may have been a hybrid machine learning and template-based code, but it still highly depends on accurate templates to produce good results. Besides, it is unable to accept morphological parameters as inputs to improve the results. A possible way is to use a machine learning code on top of a template code: ANNZ2 could use the photo-z PDFs of a template code as inputs, in addition to magnitudes and galaxy morphology. Various novel methods that involve two different codes could be explored.

4. DELIGHT *and emission lines.* We have explored on tweaking many hyperparameters and aspects of the DELIGHT algorithm to make it perform well. However, one issue when optimising DELIGHT for PAU is that we have solely depended on the Brown templates to account for emission lines in the data. It would be interesting to see how the code could be modified for manual emission line modelling, as is the case for BCNZV2.

The research in photometric redshifts remains a very relevant and essential field to pursue in the near future. As upcoming new large sky surveys like LSST and Euclid prepare in earnest for the arrival of data, we will have access to high quality galaxy photometry and morphology at fainter magnitudes than ever before, therefore photo-z estimation methods need to be constantly improved to be ready for the challenges to come. In fact, the performance of photo-z's should always be coupled directly to the relevant science cases (e.g. large-scale structures, weak lensing and galaxy clustering), this will aid the photo-z community prioritise their efforts to produce photo-z's which are best suited for each situation.

We hope that this work has been helpful to the reader, and hopefully gave new insights or perspectives on the current state of photometric redshifts. We also hope that this work would inspire new ideas to the reader, so that photometric redshifts can continue to be enhanced, in this era of precision cosmology.

BIBLIOGRAPHY

Abazajian, K. N. et al. (2004). "The Second Data Release of the Sloan Digital Sky Survey." In: *AJ* 128.1, p. 502. DOI: 10.1086/421365 (cit. on p. 109).

Abazajian, K. N. et al. (2009). "The Seventh Data Release of the Sloan Digital Sky Survey." In: *ApJS* 182.2, p. 543. DOI: 10.1088/0067-0049/182/2/543 (cit. on pp. 108, 109).

Abbott, T. et al. (2005). "The Dark Energy Survey." In: *arXiv e-prints*. arXiv: astro-ph/0510346 (cit. on pp. 82, 181).

Abdalla, F. B. et al. (2011). "A comparison of six photometric redshift methods applied to 1.5 million luminous red galaxies." In: *MNRAS* 417.3, pp. 1891–1903. DOI: 10.1111/j.1365-2966.2011.19375.x (cit. on pp. 88, 96, 123, 127, 132, 139, 162).

Abell, P. A. et al. (2009). "LSST Science Book, Version 2.0." In: *arXiv e-prints*. arXiv: 0912.0201 (cit. on p. 210).

Ade, P. a. R. et al. (2016). "Planck 2015 results - XIII. Cosmological parameters." In: *A&A* 594, A13. DOI: 10.1051/0004-6361/201525830 (cit. on pp. 40, 42, 56).

Aihara, H. et al. (2011). "The Eighth Data Release of the Sloan Digital Sky Survey: First Data from SDSS-III." In: *ApJS* 193.2, p. 29. DOI: 10.1088/0067-0049/193/2/29 (cit. on pp. 105, 108).

Aihara, H. et al. (2018). "The Hyper Suprime-Cam SSP Survey: Overview and survey design." In: *PASJ* 70 (SP1). DOI: 10.1093/pasj/psx066 (cit. on p. 82).

Alam, S. et al. (2015). "The Eleventh and Twelfth Data Releases of the Sloan Digital Sky Survey: Final Data from SDSS-III." In: *ApJS* 219.1, p. 12. DOI: 10.1088/0067-0049/219/1/12 (cit. on pp. 109, 117).

Almosallam, I. A. et al. (2016a). "A sparse Gaussian process framework for photometric redshift estimation." In: *MNRAS* 455.3, pp. 2387–2401. DOI: 10.1093/mnras/stv2425 (cit. on p. 89).

Almosallam, I. A., M. J. Jarvis, and S. J. Roberts (2016b). "GPz: non-stationary sparse Gaussian processes for heteroscedastic uncertainty estimation in photometric redshifts." In: *MNRAS* 462.1, pp. 726–739. DOI: 10.1093/mnras/stw1618 (cit. on pp. 101, 211).

Alonso, D. et al. (2017). "Calibrating photometric redshifts with intensity mapping observations." In: *Phys. Rev. D* 96.4, p. 043515. DOI: 10.1103/PhysRevD.96.043515 (cit. on p. 232).

Anderson, L. et al. (2014). "The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Releases 10 and 11 Galaxy samples." In: *MNRAS* 441.1, pp. 24–62. DOI: 10.1093/mnras/stu523 (cit. on p. 56).

Annis, J. et al. (2014). "The Sloan Digital Sky Survey Coadd: 275 deg2 of Deep Sloan Digital Sky Survey Imaging on Stripe 82." In: *ApJ* 794.2, p. 120. DOI: 10.1088/0004-637X/794/2/120 (cit. on pp. 108, 109, 184).

Antolini, E. and J. S. Heyl (2016). "Using the 2MASS photometric redshift survey to optimize LIGO follow-up observations." In: *MNRAS* 462.1, pp. 1085–1091. DOI: 10.1093/mnras/stw1720 (cit. on p. 82).

Arnouts, S. et al. (1999). "Measuring and modelling the redshift evolution of clustering: the Hubble Deep Field North." In: *MNRAS* 310.2, pp. 540–556. DOI: 10.1046/j.1365-8711.1999.02978.x (cit. on pp. 85, 91, 101, 139, 211).

Arnouts, S. et al. (2002). "Measuring the redshift evolution of clustering: the Hubble Deep Field South." In: *MNRAS* 329.2, pp. 355–366. DOI: 10.1046/j.1365-8711.2002.04988.x (cit. on p. 234).

Ash, M. E., I. I. Shapiro, and W. B. Smith (1967). "Astronomical constants and planetary ephemerides deduced from radar and optical observations." In: *AJ* 72, p. 338. DOI: 10.1086/110230 (cit. on p. 44).

Assef, R. J. et al. (2008). "Low-Resolution Spectral Templates for Galaxies from 0.2 to 10 mu m." In: *ApJ* 676.1, p. 286. DOI: 10.1086/527533 (cit. on pp. 90, 101).

Babbedge, T. S. R. et al. (2004). "ImpZ: a new photometric redshift code for galaxies and quasars." In: *MNRAS* 353.2, pp. 654–672. DOI: 10.1111/j.1365-2966.2004.08105.x (cit. on pp. 90, 101).

Ball, N. M. et al. (2004). "Galaxy types in the Sloan Digital Sky Survey using supervised artificial neural networks." In: *MNRAS* 348.3, pp. 1038–1046. DOI: 10.1111/j.1365-2966.2004.07429.x (cit. on p. 123).

Ball, N. M. et al. (2007). "Robust Machine Learning Applied to Astronomical Data Sets. II. Quantifying Photometric Redshifts for Quasars Using Instance-based Learning." In: *ApJ* 663.2, p. 774. DOI: 10.1086/518362 (cit. on p. 89).

Ball, N. M. et al. (2008). "Robust Machine Learning Applied to Astronomical Data Sets. III. Probabilistic Photometric Redshifts for Galaxies and Quasars in the SDSS and GALEX." In: *ApJ* 683.1, p. 12. DOI: 10.1086/589646 (cit. on p. 92).

Banerji, M. et al. (2010). "Galaxy Zoo: reproducing galaxy morphologies via machine learning." In: *MNRAS* 406.1, pp. 342–353. DOI: 10.1111/j.1365-2966.2010.16713.x (cit. on pp. 69, 108, 123, 151).

Barish, B. C. and R. Weiss (2008). "LIGO and the Detection of Gravitational Waves." In: *Physics Today* 52.10, p. 44. DOI: 10.1063/1.882861 (cit. on p. 82).

Barro, G. et al. (2011). "UV-to-FIR Analysis of Spitzer/IRAC Sources in the Extended Groth Strip. II. Photometric Redshifts, Stellar Masses, and Star Formation Rates." In: *ApJS* 193.2, p. 30. DOI: 10.1088/0067-0049/193/2/30 (cit. on p. 101).

Batista, G. E. A. P. A. and M. C. Monard (2003). "An analysis of four missing data treatment methods for supervised learning." In: *Applied Artificial Intelligence* 17.5, pp. 519–533. DOI: 10.1080/713827181 (cit. on p. 212).

Battaglia, N. et al. (2016). "Weak-lensing mass calibration of the Atacama Cosmology Telescope equatorial Sunyaev-Zeldovich cluster sample with the Canada-France-Hawaii telescope stripe 82 survey." In: *Journal of Cosmology and Astroparticle Physics* 2016.8, p. 013. DOI: 10.1088/1475-7516/2016/08/013 (cit. on p. 111).

Baum, W. A. (1962). "Photoelectric Magnitudes and Red-Shifts." In: *Problems of Extra-Galactic Research*. Ed. by G. C. McVittie. Vol. 15. IAU Symposium. New York, USA: Macmillan Press, p. 390. URL: http://adsabs.harvard.edu/abs/1962IAUS...15..390B (cit. on p. 82).

Beck, R. et al. (2017). "Photo-z-SQL: Integrated, flexible photometric redshift computation in a database." In: *Astron. Comput.* 19, pp. 34–44. DOI: 10.1016/j.ascom.2017.03.002 (cit. on p. 101).

Becker, R. H., R. L. White, and D. J. Helfand (1995). "The FIRST Survey: Faint Images of the Radio Sky at Twenty Centimeters." In: *ApJ* 450, p. 559. DOI: 10.1086/176166 (cit. on p. 110).

Behroozi, P. S., R. H. Wechsler, and H-Y. Wu (2013). "The ROCKSTAR Phase-space Temporal Halo Finder and the Velocity Offsets of Cluster Cores." In: *ApJ* 762.2, p. 109. DOI: 10.1088/0004-637X/762/2/109 (cit. on p. 112).

Bekenstein, J. D. (2004). "Relativistic gravitation theory for the modified Newtonian dynamics paradigm." In: *Phys. Rev. D* 70.8, p. 083509. DOI: 10.1103/PhysRevD.70.083509 (cit. on p. 41).

Bender, R. et al. (2001). "The FORS Deep Field: Photometric Data and Photometric Redshifts." In: *Deep Fields*. ESO Astrophysics Symposia. Ed. by S. Cristiani, A. Renzini, and R. E. Williams. Proceedings of The ESO Workshop. Garching, Germany: Springer, Berlin, Heidelberg, pp. 96–101. ISBN: 978-3-540-45514-1. DOI: 10.1007/10854354_18 (cit. on pp. 96, 101).

Benítez, N. (2000). "Bayesian Photometric Redshift Estimation." In: *ApJ* 536.2, p. 571. DOI: 10.1086/308947 (cit. on pp. 85, 90–92, 101, 132, 133, 192, 211).

Benjamin, J. et al. (2013). "CFHTLenS tomographic weak lensing: quantifying accurate redshift distributions." In: *MNRAS* 431.2, pp. 1547–1564. DOI: 10.1093/mnras/stt276 (cit. on p. 82).

Benn, C. R and S. L Ellison (1998). "Brightness of the night sky over La Palma." In: *New Astronomy Reviews* 42.6, pp. 503–507. DOI: 10.1016/S1387-6473(98)00062-1 (cit. on p. 114).

Bernstein, G. M. and M. J. Jarvis (2002). "Shapes and Shears, Stars and Smears: Optimal Measurements for Weak Lensing." In: *AJ* 123.2, p. 583. DOI: 10.1086/338085 (cit. on p. 71).

Berry, M. J. A. and G. S. Linoff (2004). *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*. John Wiley & Sons. 671 pp. ISBN: 978-0-471-47064-9 (cit. on p. 125).

Bertin, E. and S. Arnouts (1996). "SExtractor: Software for source extraction." In: *A&AS* 117.2, pp. 393–404. DOI: 10.1051/aas:1996164 (cit. on pp. 68, 111, 123).

Bessel, F. W. (1839). "Bestimmung der Entfernung des 61sten Sterns des Schwans." In: *Astron. Nachr.* 16.5, pp. 65–96. DOI: 10.1002/asna.18390160502 (cit. on p. 45).

Bessell, M. S. (2005). "Standard Photometric Systems." In: *Annual Review of Astronomy and Astrophysics* 43.1, pp. 293–336. DOI: 10.1146/annurev.astro.41.082801.100251 (cit. on pp. 47, 61, 62).

Bianchini, F. and C. L. Reichardt (2018). "Constraining gravity at large scales with the 2MASS Photometric Redshift catalogue and Planck lensing." In: *arXiv e-prints*. arXiv: 1801.03736 (cit. on p. 82).

Bigelow, B. C. and A. M. Dressler (2003). "IMACS, the multiobject spectrograph and imager for Magellan: a status report." In: *Instrument Design and Performance for Optical/Infrared Ground-based Telescopes*. Astronomical Telescopes and Instrumentation. Vol. 4841. Proc. SPIE. Waikoloa, Hawaii, USA: International Society for Optics and Photonics, pp. 1727–1739. DOI: 10.1117/12.461870 (cit. on p. 121).

Bilicki, M. et al. (2017). "Photometric redshifts for the Kilo-Degree Survey. Machine-learning analysis with artificial neural networks." In: *arXiv e-prints*. arXiv: 1709.04205 (cit. on p. 227).

Blanton, M. R. and S. Roweis (2007). "K-Corrections and Filter Transformations in the Ultraviolet, Optical, and Near-Infrared." In: *AJ* 133.2, p. 734. DOI: 10.1086/510127 (cit. on pp. 65, 101, 141).

Blanton, M. R. et al. (2001). "The Luminosity Function of Galaxies in SDSS Commissioning Data." In: *AJ* 121.5, p. 2358. DOI: 10.1086/320405 (cit. on p. 67).

Blanton, M. R. et al. (2005). "New York University Value-Added Galaxy Catalog: A Galaxy Catalog Based on New Public Surveys." In: *AJ* 129.6, p. 2562. DOI: 10.1086/429803 (cit. on pp. 68, 72, 109).

Blum, A. (1992). *Neural Networks in C++: An Object-Oriented Framework for Building Connectionist Systems*. John Wiley & Sons. 228 pp. ISBN: 978-0-471-55201-7 (cit. on pp. 125, 127).

Bolton, A. S. et al. (2012). "Spectral Classification and Redshift Measurement for the SDSS-III Baryon Oscillation Spectroscopic Survey." In: *AJ* 144.5, p. 144. DOI: 10.1088/0004-6256/144/5/144 (cit. on p. 76).

Bolzonella, M., J-M. Miralles, and R. Pelló (2000). "Photometric redshifts based on standard SED fitting procedures." In: *A&A* 363.2, pp. 476–492. URL: http://aa.springer.de/papers/0363002/2300476.pdf (cit. on pp. 85, 101).

Bonanos, A. Z. (2006). "Eclipsing Binaries: Tools for Calibrating the Extragalactic Distance Scale." In: *Proc. IAU* 2 (S240), pp. 79–87. DOI: 10.1017/S1743921307003845 (cit. on p. 56).

Bonfield, D. G. et al. (2010). "Photometric redshift estimation using Gaussian processes." In: *MNRAS* 405.2, pp. 987–994. DOI: 10.1111/j.1365-2966.2010.16544.x (cit. on pp. 95, 123, 127, 243).

Bonnett, C. (2015). "Using neural networks to estimate redshift distributions. An application to CFHTLenS." In: *MNRAS* 449.1, pp. 1043–1056. DOI: 10.1093/mnras/stv230 (cit. on pp. 93, 128).

Bonnett, C. et al. (2016). "Redshift distributions of galaxies in the Dark Energy Survey Science Verification shear catalogue and implications for weak lensing." In: *Phys. Rev. D* 94.4, p. 042005. DOI: 10.1103/PhysRevD.94.042005 (cit. on pp. 96, 121).

Boulade, O. et al. (2003). "MegaCam: the new Canada-France-Hawaii Telescope wide-field imaging camera." In: *Instrument Design and Performance for Optical/Infrared Ground-based Telescopes*. Astronomical Telescopes and Instrumentation. Vol. 4841. Proc. SPIE. Waikoloa, Hawaii, USA: International Society for Optics and Photonics, pp. 72–82. DOI: 10.1117/12.459890 (cit. on pp. 105, 110, 112).

Bramich, D. M. et al. (2008). "Light and motion in SDSS Stripe 82: the catalogues." In: *MNRAS* 386.2, pp. 887–902. DOI: 10.1111/j.1365-2966.2008.13053.x (cit. on pp. 109, 184).

Brammer, G. B., P. G. Van Dokkum, and P. Coppi (2008). "EAZY: A Fast, Public Photometric Redshift Code." In: *ApJ* 686.2, p. 1503. DOI: 10.1086/591786 (cit. on pp. 85, 101, 173, 211).

Breiman, L. (1996). "Bagging predictors." In: *Machine Learning* 24.2, pp. 123–140. DOI: 10.1007/BF00058655 (cit. on p. 129).

Breiman, L. et al. (1984). *Classification and Regression Trees*. Taylor and Francis. 372 pp. ISBN: 978-0-412-04841-8 (cit. on p. 128).

Brescia, M. et al. (2014). "A catalogue of photometric redshifts for the SDSS-DR9 galaxies." In: *A&A* 568, A126. DOI: 10.1051/0004-6361/201424383 (cit. on pp. 96, 101).

Brescia, M. et al. (2018). "The astronomical data deluge and the template case of photometric redshifts." In: *arXiv e-prints*. arXiv: 1802.07683 (cit. on p. 99).

Brown, A. G. A. et al. (2018). "Gaia Data Release 2. Summary of the contents and survey properties." In: *arXiv e-prints*. arXiv: 1804.09365. URL: http://arxiv.org/abs/1804.09365 (cit. on p. 46).

Brown, M. J. I. et al. (2014). "An Atlas of Galaxy Spectral Energy Distributions from the Ultraviolet to the Mid-infrared." In: *ApJS* 212.2, p. 18. DOI: 10.1088/0067-0049/212/2/18 (cit. on pp. 91, 139, 140, 236).

Brun, R. and F. Rademakers (1997). "ROOT - An object oriented data analysis framework." In: *Nuclear Instruments and Methods in Physics Research Section A* 389.1, pp. 81–86. DOI: 10.1016/S0168-9002(97)00048-X (cit. on p. 127).

Brunner, R. J. et al. (1997). "Toward More Precise Photometric Redshifts: Calibration via CCD Photometry." In: *ApJ* 482.1, p. L21. DOI: 10.1086/310674 (cit. on p. 85).

Bruzual, A. G. (1983). "Spectral evolution of galaxies. I - Early-type systems." In: *ApJ* 273, pp. 105–127. DOI: 10.1086/161352 (cit. on pp. 83, 84).

Bruzual, A. G. and S. Charlot (2003). "Stellar population synthesis at the resolution of 2003." In: *MNRAS* 344.4, pp. 1000–1028. DOI: 10.1046/j.1365-8711.2003.06897.x (cit. on pp. 90, 139, 140, 234, 236).

Budavári, T. (2009). "A Unified Framework for Photometric Redshifts." In: *ApJ* 695.1, p. 747. DOI: 10.1088/0004-637X/695/1/747 (cit. on p. 90).

Bundy, K. et al. (2012). "SYNMAG Photometry: A Fast Tool for Catalog-level Matched Colors of Extended Sources." In: *AJ* 144.6, p. 188. DOI: 10.1088/0004-6256/144/6/188 (cit. on pp. 105, 110, 173).

Bundy, K. et al. (2015). "The Stripe 82 Massive Galaxy Project. I. Catalog Construction." In: *ApJS* 221.1, p. 15. DOI: 10.1088/0067-0049/221/1/15 (cit. on pp. 70, 109, 110, 123, 132, 168, 173, 175).

Caballero, J. A. (2008). "Dynamical parallax of sigma Ori AB: mass, distance and age." In: *MNRAS* 383.2, pp. 750–754. DOI: 10.1111/j.1365-2966.2007.12614.x (cit. on p. 57).

Calvi, V. et al. (2014). "The Effect of Surface Brightness Dimming in the Selection of High-z Galaxies." In: *ApJ* 796.2, p. 102. DOI: 10.1088/0004-637X/796/2/102 (cit. on p. 72).

Calzetti, D., A. L. Kinney, and T. Storchi-Bergmann (1994). "Dust extinction of the stellar continua in starburst galaxies: The ultraviolet and optical extinction law." In: *ApJ* 429, pp. 582–601. DOI: 10.1086/174346 (cit. on pp. 139, 236).

Calzetti, D. et al. (2000). "The Dust Content and Opacity of Actively Star-forming Galaxies." In: *ApJ* 533.2, p. 682. DOI: 10.1086/308692 (cit. on p. 140).

Cannon, R. et al. (2006). "The 2dF-SDSS LRG and QSO (2SLAQ) Luminous Red Galaxy Survey." In: *MNRAS* 372.1, pp. 425–442. DOI: 10.1111/j.1365-2966.2006.10875.x (cit. on pp. 118, 161).

Cao, Y. et al. (2017). "Dependency of photo-z accuracy on filter definition of photometric surveys." In: *arXiv e-prints*. arXiv: 1706.09586 (cit. on p. 98).

Carliles, S. et al. (2010). "Random Forests for Photometric Redshifts." In: *ApJ* 712.1, p. 511. DOI: 10.1088/0004-637X/712/1/511 (cit. on pp. 89, 94).

Carrasco Kind, M. and R. J. Brunner (2013). "TPZ: photometric redshift PDFs and ancillary information by using prediction trees and random forests." In: *MNRAS* 432.2, pp. 1483–1501. DOI: 10.1093/mnras/stt574 (cit. on pp. 96, 102, 211).

— (2014a). "SOMz: photometric redshift PDFs with self-organizing maps and random atlas." In: *MNRAS* 438.4, pp. 3409–3421. DOI: 10.1093/mnras/stt2456 (cit. on pp. 97, 98, 102).

Carrasco Kind, M. and R. J. Brunner (2014b). "Sparse representation of photometric redshift probability density functions: preparing for petascale astronomy." In: *MNRAS* 441.4, pp. 3550–3561. DOI: 10.1093/mnras/stu827 (cit. on p. 93).

Carretero, J. et al. (2015). "An algorithm to build mock galaxy catalogues using MICE simulations." In: *MNRAS* 447.1, pp. 646–670. DOI: 10.1093/mnras/stu2402 (cit. on p. 114).

Carretero, J. et al. (2018). "CosmoHub and SciPIC: Massive cosmological data analysis, distribution and generation using a Big Data platform." In: *Proceedings of The European Physical Society Conference on High Energy Physics*. The European Physical Society Conference on High Energy Physics. Vol. 314. Proceedings of The European Physical Society Conference on High Energy Physics. SISSA Medialab, p. 488. DOI: 10.22323/1.314.0488 (cit. on p. 114).

Castander, F. J. et al. (2012). "The PAU camera and the PAU survey at the William Herschel Telescope." In: *Ground-based and Airborne Instrumentation for Astronomy IV*. SPIE Astronomical Telescopes + Instrumentation. Vol. 8446. Proc. SPIE. Amsterdam, Netherlands: International Society for Optics and Photonics, p. 84466D. DOI: 10.1117/12.926234 (cit. on pp. 82, 111).

Castignani, G. and C. Benoist (2016). "A new method to assign galaxy cluster membership using photometric redshifts." In: *A&A* 595, A111. DOI: 10.1051/0004-6361/201528009 (cit. on p. 82).

Castignani, G. et al. (2014). "A New Method to Search for High-redshift Clusters Using Photometric Redshifts." In: *ApJ* 792.2, p. 113. DOI: 10.1088/0004-637X/792/2/113 (cit. on p. 82).

Cavuoti, S. et al. (2015). "Photometric redshift estimation based on data mining with PhotoRApToR." In: *Exp. Astron.* 39.1, pp. 45–71. DOI: 10.1007/s10686-015-9443-4 (cit. on pp. 98, 101).

Cavuoti, S. et al. (2017a). "A cooperative approach among methods for photometric redshifts estimation: an application to KiDS data." In: *MNRAS* 466.2, pp. 2039–2053. DOI: 10.1093/mnras/stw3208 (cit. on pp. 90, 96, 132).

Cavuoti, S. et al. (2017b). "METAPHOR: a machine-learning-based method for the probability density estimation of photometric redshifts." In: *MNRAS* 465.2, pp. 1959–1973. DOI: 10.1093/mnras/stw2930 (cit. on pp. 98, 101, 211).

Charbonnier, A. et al. (2017). "The abundance of compact quiescent galaxies since z about 0.6." In: *MNRAS* 469.4, pp. 4523–4536. DOI: 10.1093/mnras/stx1142 (cit. on p. 111).

Chevallard, J. and S. Charlot (2016). "Modelling and interpreting spectral energy distributions of galaxies with beagle." In: *MNRAS* 462.2, pp. 1415–1443. DOI: 10.1093/mnras/stw1756 (cit. on p. 101).

Chiappetti, L. et al. (2005). "The XMM-LSS survey - The XMDS/VVDS 4 sigma catalogue." In: *A&A* 439.1, pp. 413–425. DOI: 10.1051/0004-6361:20042583 (cit. on p. 140).

Chilingarian, I. V., A.-L. Melchior, and I. V. Zolotukhin (2010). "Analytical approximations of K-corrections in optical and near-infrared bands." In: *MNRAS* 405.3, pp. 1409–1420. DOI: 10.1111/j.1365-2966.2010.16506.x (cit. on p. 65).

Chromey, F. R. (2010). *To Measure the Sky: An Introduction to Observational Astronomy*. Cambridge University Press. 461 pp. ISBN: 978-0-521-76386-8 (cit. on p. 63).

Claussen, M. (2004). "Astronomical Masers." In: *Science* 306.5694, pp. 235–236. DOI: 10.1126/science.1101353 (cit. on p. 49).

Clementini, G. et al. (2017). "Gaia Data Release 1 - Testing parallaxes with local Cepheids and RR Lyrae stars." In: *A&A* 605, A79. DOI: 10.1051/0004-6361/201629925 (cit. on p. 46).

Coe, D. et al. (2006). "Galaxies in the Hubble Ultra Deep Field. I. Detection, Multiband Photometry, Photometric Redshifts, and Morphology." In: *AJ* 132.2, p. 926. DOI: 10.1086/505530 (cit. on pp. 138, 139).

Coil, A. L. et al. (2011). "The PRIsm MUlti-object Survey (PRIMUS). I. Survey Overview and Characteristics." In: *ApJ* 741.1, p. 8. DOI: 10.1088/0004-637X/741/1/8 (cit. on pp. 81, 120).

Cole, S. et al. (2001). "The 2dF galaxy redshift survey: near-infrared galaxy luminosity functions." In: *MNRAS* 326.1, pp. 255–273. DOI: 10.1046/j.1365-8711.2001.04591.x (cit. on p. 65).

Coleman, G. D., C.-C. Wu, and D. W. Weedman (1980). "Colors and magnitudes predicted for high redshift galaxies." In: *ApJS* 43, pp. 393–416. DOI: 10.1086/190674 (cit. on pp. 90, 138–140).

Colless, M. et al. (2001). "The 2dF Galaxy Redshift Survey: spectra and redshifts." In: *MNRAS* 328.4, pp. 1039–1063. DOI: 10.1046/j.1365-8711.2001.04902.x (cit. on p. 118).

Collister, A. A. and O. Lahav (2004). "ANNz: estimating photometric redshifts using artificial neural networks." In: *PASP* 116.818, pp. 345–351. DOI: 10.1086/383254 (cit. on pp. 86, 88, 97, 101, 123, 125, 127, 243).

Collister, A. A. et al. (2007). "MegaZ-LRG: a photometric redshift catalogue of one million SDSS luminous red galaxies." In: *MNRAS* 375.1, pp. 68–76. DOI: 10.1111/j.1365-2966.2006.11305.x (cit. on pp. 86, 123, 127).

Comparat, J. et al. (2013). "Stochastic bias of colour-selected BAO tracers by joint clustering-weak lensing analysis." In: *MNRAS* 433.2, pp. 1146–1160. DOI: 10.1093/mnras/stt797 (cit. on p. 111).

Connolly, A. J. et al. (1995). "Slicing Through Multicolor Space: Galaxy Redshifts from Broadband Photometry." In: *AJ* 110, p. 2655. DOI: 10.1086/117720 (cit. on p. 85).

Copeland, E. J., M. Sami, and S. Tsujikawa (2006). "Dynamics of dark energy." In: *International Journal of Modern Physics* 15.11, pp. 1753–1935. DOI: 10.1142/S021827180600942X (cit. on p. 52).

Cousins, A. W. J. (1980). "Fainter Standards for VRI Photometry in the E Regions." In: *Monthly Notes of the Astronomical Society of South Africa* 39, p. 22. URL: https://ui.adsabs.harvard.edu/#abs/1980MNSSA..39...22C/abstract (cit. on p. 61).

Craig, M., J. Cabanela, and L. Winkler (2014). *Introduction to Astronomical Image Analaysis*. URL: http://image-analysis.readthedocs.io/en/latest/index.html# (cit. on p. 62).

Crocce, M. et al. (2010). "Simulating the Universe with MICE: the abundance of massive clusters." In: *MNRAS* 403.3, pp. 1353–1367. DOI: 10.1111/j.1365-2966.2009.16194.x (cit. on p. 114).

Crocce, M. et al. (2015). "The MICE Grand Challenge lightcone simulation - II. Halo and galaxy catalogues." In: *MNRAS* 453.2, pp. 1513–1530. DOI: 10.1093/mnras/stv1708 (cit. on p. 114).

Croom, S. M. et al. (2009). "The 2dF-SDSS LRG and QSO Survey: the spectroscopic QSO catalogue." In: *MNRAS* 392.1, pp. 19–44. DOI: 10.1111/j.1365-2966.2008.14052.x (cit. on p. 116).

Cunha, C. E. et al. (2009). "Estimating the redshift distribution of photometric galaxy samples - II. Applications and tests of a new method." In: *MNRAS* 396.4, pp. 2379–2398. DOI: 10.1111/j.1365-2966.2009.14908.x (cit. on p. 95).

D'Isanto, A. and K. L. Polsterer (2018). "Photometric redshift estimation via deep learning - Generalized and pre-classification-less, image based, fully probabilistic redshifts." In: *A&A* 609, A111. DOI: 10.1051/0004-6361/201731326 (cit. on p. 99).

D'Isanto, A. et al. (2018). "Return of the features - Efficient feature selection and interpretation for photometric redshifts." In: *arXiv e-prints*. arXiv: 1803.10032 (cit. on p. 98).

Daddi, E. et al. (2004). "A New Photometric Technique for the Joint Selection of Star-forming and Passive Galaxies at 1.4<z<2.5." In: *ApJ* 617.2, p. 746. DOI: 10.1086/425569 (cit. on p. 120).

Dahlen, T. et al. (2005). "The Evolution of the Optical and Near-Infrared Galaxy Luminosity Functions and Luminosity Densities to z about 2." In: *ApJ* 631.1, p. 126. DOI: 10.1086/432027 (cit. on pp. 60, 114).

Dahlen, T. et al. (2010). "A Detailed Study of Photometric Redshifts for GOODS-South Galaxies." In: *ApJ* 724.1, p. 425. DOI: 10.1088/0004-637X/724/1/425 (cit. on pp. 91, 101).

Dahlen, T. et al. (2013). "A Critical Assesment of Photometric Redshift Methods: A Candels Investigation." In: *ApJ* 775.2, p. 93. DOI: 10.1088/0004-637X/775/2/93 (cit. on pp. 90, 93, 94, 96).

Davis, M. and D. W. Latham (1979). "Photon-Counting Reticon Detector." In: *Instrumentation in Astronomy III*. Instrumentation in Astronomy. Vol. 0172. Tucson, USA: International Society for Optics and Photonics, pp. 71–82. DOI: 10.1117/12.957069 (cit. on p. 78).

Dawson, K. S. et al. (2013). "The Baryon Oscillation Spectroscopic Survey of SDSS-III." In: *AJ* 145.1, p. 10. DOI: 10.1088/0004-6256/145/1/10 (cit. on pp. 117, 118).

De Vicente, J., E. Sanchez, and I. Sevilla-Noarbe (2016). "DNF - Galaxy photometric redshift by Directional Neighbourhood Fitting." In: *MNRAS* 459.3, pp. 3078–3088. DOI: 10.1093/mnras/stw857 (cit. on p. 89).

Díaz Tello, J. et al. (2016). "Star formation activity in Balmer break galaxies at z<1.5." In: *A&A* 587, A136. DOI: 10.1051/0004-6361/201526315 (cit. on p. 75).

Djorgovski, S. and M. Davis (1987). "Fundamental properties of elliptical galaxies." In: *ApJ* 313, pp. 59–68. DOI: 10.1086/164948 (cit. on p. 53).

Dressler, A. et al. (1987). "Spectroscopy and photometry of elliptical galaxies. I - A new distance estimator." In: *ApJ* 313, pp. 42–58. DOI: 10.1086/164947 (cit. on p. 53).

Drinkwater, M. J. et al. (2010). "The WiggleZ Dark Energy Survey: survey design and first data release." In: *MNRAS* 401.3, pp. 1429–1452. DOI: 10.1111/j.1365-2966.2009.15754.x (cit. on pp. 76, 119, 120).

Drucker, H. (1997). "Improving Regressors Using Boosting Techniques." In: *Proceedings of the Fourteenth International Conference on Machine Learning*. Proc. International Conference on Machine Learning 14. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., pp. 107–115. ISBN: 978-1-55860-486-5 (cit. on p. 129).

Duncan, K. J. et al. (2017). "Photometric redshifts for the next generation of deep radio continuum surveys - II. Gaussian processes and hybrid estimates." In: *arXiv e-prints*. arXiv: 1712.04476 (cit. on p. 96).

Duncan, K. J. et al. (2018). "Photometric redshifts for the next generation of deep radio continuum surveys - I. Template fitting." In: *MNRAS* 473.2, pp. 2655–2672. DOI: 10.1093/mnras/stx2536 (cit. on p. 96).

Dyson, F. W. (1913). "The Distribution in Space of the Stars in Carrington's Circumpolar Catalogue." In: *MNRAS* 73.5, pp. 334–342. DOI: 10.1093/mnras/73.5.334 (cit. on p. 45).

Ebden, M. (2015). "Gaussian Processes: A Quick Introduction." In: *arXiv e-prints*. arXiv: 1505.02965 (cit. on p. 135).

Eddington, A. S. (1920). "The Internal Constitution of the Stars." In: *The Scientific Monthly* 11.4, pp. 297–303. URL: http://www.jstor.org/stable/6491 (cit. on p. 48).

Edge, A. et al. (2013). "The VISTA Kilo-degree Infrared Galaxy (VIKING) Survey: Bridging the Gap between Low and High Redshift." In: *The Messenger* 154, pp. 32–34. URL: http://adsabs.harvard.edu/abs/2013Msngr.154...32E (cit. on p. 227).

Einstein, A. (1916). "Die Grundlage der allgemeinen Relativitatstheorie." In: *Ann. Phys. (Berl.)* 354.7, pp. 769–822. DOI: 10.1002/andp.19163540702 (cit. on p. 40).

Eisenstein, D. J. et al. (2001). "Spectroscopic Target Selection for the Sloan Digital Sky Survey: The Luminous Red Galaxy Sample." In: *AJ* 122.5, p. 2267. DOI: 10.1086/323717 (cit. on pp. 115, 116, 118).

Eisenstein, D. J. et al. (2005). "Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies." In: *ApJ* 633.2, p. 560. DOI: 10.1086/466512 (cit. on pp. 55, 56, 116).

Eisenstein, D. J. et al. (2011). "SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way, and Extra-Solar Planetary Systems." In: *AJ* 142.3, p. 72. DOI: 10.1088/0004-6256/142/3/72 (cit. on p. 117).

Elliott, J. et al. (2015). "The overlooked potential of Generalized Linear Models in astronomy-II: Gamma regression and photometric redshifts." In: *Astron. Comput.* 10, pp. 61–72. DOI: 10.1016/j.ascom.2015.01.002 (cit. on p. 89).

Ewen, H. I. and E. M. Purcell (1951). "Observation of a Line in the Galactic Radio Spectrum: Radiation from Galactic Hydrogen at 1,420 Mc. per sec." In: *Nature* 168.4270, p. 356. DOI: 10.1038/168356a0 (cit. on p. 75).

Faber, S. M. and R. E. Jackson (1976). "Velocity dispersions and mass-to-light ratios for elliptical galaxies." In: *ApJ* 204, pp. 668–683. DOI: 10.1086/154215 (cit. on p. 52).

Faber, S. M. et al. (2003). "The DEIMOS spectrograph for the Keck II Telescope: integration and testing." In: *Instrument Design and Performance for Optical/Infrared Ground-based Telescopes*. Astronomical Telescopes and Instrumentation. Vol. 4841. Proc. SPIE. Waikoloa, Hawaii, USA: International Society for Optics and Photonics, pp. 1657–1670. DOI: 10.1117/12.460346 (cit. on p. 118).

Feldmann, R. et al. (2006). "The Zurich Extragalactic Bayesian Redshift Analyzer and its first application: COSMOS." In: *MNRAS* 372.2, pp. 565–577. DOI: 10.1111/j.1365-2966.2006.10930.x (cit. on pp. 85, 101).

Felten, J. E. (1976). "On Schmidt's Vm estimator and other estimators of luminosity functions." In: *ApJ* 207, p. 700. DOI: 10.1086/154538 (cit. on p. 237).

Fernández-Soto, A. et al. (2002). "Error analysis of the photometric redshift technique." In: *MNRAS* 330.4, pp. 889–894. DOI: 10.1046/j.1365-8711.2002.05131.x (cit. on pp. 92, 94).

Finlator, K., R. Dave, and B. D. Oppenheimer (2007). "Constraints on physical properties of z about 6 galaxies using cosmological hydrodynamic simulations." In: *MNRAS* 376.4, pp. 1861–1878. DOI: 10.1111/j.1365-2966.2007.11578.x (cit. on p. 101).

Firth, A. E., O. Lahav, and R. S. Somerville (2003). "Estimating photometric redshifts with artificial neural networks." In: *MNRAS* 339.4, pp. 1195–1202. DOI: 10.1046/j.1365-8711.2003.06271.x (cit. on pp. 86, 89).

Fletcher, R. (1987). *Practical methods of optimization*. Wiley. 462 pp. ISBN: 978-0-471-91547-8 (cit. on p. 126).

Fliri, J. and I. Trujillo (2016). "The IAC Stripe 82 Legacy Project: a wide-area survey for faint surface brightness astronomy." In: *MNRAS* 456.2, pp. 1359–1373. DOI: 10.1093/mnras/stv2686 (cit. on p. 109).

Fosalba, P. et al. (2015). "The MICE grand challenge lightcone simulation - I. Dark matter clustering." In: *MNRAS* 448.4, pp. 2987–3000. DOI: 10.1093/mnras/stv138 (cit. on p. 114).

Fraunhofer, J. (1814). "Bestimmung des Brechungs - und des Farben-Zerstreuungs - Vermogens verschiedener Glasarten, in Bezug auf die Vervollkommnung achromatischer Fernrohre." In: *Denkschriften der Koniglichen Akademie der Wissenschaften zu Munchen* 5, pp. 193–226 (cit. on p. 73).

Freedman, W. L. (2017). "Cosmology at a crossroads." In: *Nature Astronomy* 1, p. 0121. DOI: 10.1038/s41550-017-0121 (cit. on p. 42).

Freeman, K. C. (1970). "On the Disks of Spiral and So Galaxies." In: *ApJ* 160, p. 811. DOI: 10.1086/150474 (cit. on p. 66).

Freeman, P. E. et al. (2009). "Photometric redshift estimation using spectral connectivity analysis." In: *MNRAS* 398.4, pp. 2012–2021. DOI: 10.1111/j.1365-2966.2009.15236.x (cit. on p. 89).

Freund, Y. and R. E. Schapire (1997). "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting." In: *Journal of Computer and System Sciences* 55.1, pp. 119–139. DOI: 10.1006/jcss.1997.1504 (cit. on p. 129).

Friedman, A. (1922). "Uber die Krummung des Raumes." In: *Zeitschrift fur Physik* 10.1, pp. 377–386. DOI: 10.1007/BF01332580 (cit. on pp. 39, 50).

Frieman, J. A. et al. (2008). "The Sloan Digital Sky Survey-II Supernova Survey: Technical Summary." In: *AJ* 135.1, p. 338. DOI: 10.1088/0004-6256/135/1/338 (cit. on pp. 109, 184).

Fritz Benedict, G. et al. (2007). "Hubble Space Telescope Fine Guidance Sensor Parallaxes of Galactic Cepheid Variable Stars: Period-Luminosity Relations." In: *AJ* 133.4, p. 1810. DOI: 10.1086/511980 (cit. on p. 49).

Fukugita, M. et al. (1996). "The Sloan Digital Sky Survey Photometric System." In: *AJ* 111, p. 1748. DOI: 10.1086/117915 (cit. on pp. 61, 78, 105).

Garilli, B. et al. (2008). "The Vimos VLT deep survey." In: *A&A* 486.3, pp. 683–695. DOI: 10.1051/0004-6361:20078878 (cit. on p. 119).

Geller, M. J. and J. P. Huchra (1989). "Mapping the Universe." In: *Science* 246.4932, pp. 897–903. DOI: 10.1126/science.246.4932.897 (cit. on p. 78).

Gendre, B., A. Galli, and M. Boer (2008). "X-ray afterglow light curves: toward a standard candle?" In: *AIP Conference Proceedings* 1000.1, pp. 72–75. DOI: 10.1063/1.2943553 (cit. on p. 57).

Gerdes, D. W. et al. (2010). "ArborZ: photometric redshifts using boosted decision trees." In: *ApJ* 715.2, p. 823. DOI: 10.1088/0004-637X/715/2/823 (cit. on pp. 89, 93, 96, 101, 128).

Giallongo, E. et al. (1998). "The Photometric Redshift Distribution and Evolutionary Properties of Galaxies up to z about 4.5 in the Field of the Quasar BR 1202-0725." In: *AJ* 115.6, p. 2169. DOI: 10.1086/300361 (cit. on p. 101).

Giavalisco, M. et al. (2004). "The Great Observatories Origins Deep Survey: Initial Results from Optical and Near-Infrared Imaging." In: *ApJ* 600.2, p. L93. DOI: 10.1086/379232 (cit. on p. 96).

Gilbank, D. G. et al. (2011). "The Red-sequence Cluster Survey-2 (RCS-2): Survey Details and Photometric Catalog Construction." In: *AJ* 141.3, p. 94. DOI: 10.1088/0004-6256/141/3/94 (cit. on pp. 120, 181).

Gillis, B. R. and M. J. Hudson (2011). "Group-finding with photometric redshifts: the photo-z probability peaks algorithm." In: *MNRAS* 410.1, pp. 13–26. DOI: 10.1111/j.1365-2966.2010.17415.x (cit. on p. 82).

Gomes, Z. et al. (2018). "Improving photometric redshift estimation using GPz: size information, post processing, and improved photometry." In: *MNRAS* 475.1, pp. 331–342. DOI: 10.1093/mnras/stx3187 (cit. on pp. 98, 145).

Goodricke, John and Bayer (1786). "A Series of Observations on, and a Discovery of, the Period of the Variation of the Light of the Star Marked delta by Bayer, Near the Head of Cepheus. In a Letter from John Goodricke, Esq. to Nevil Maskelyne, D. D. F. R. S. and Astronomer Royal." In: *Phil. Trans. R. Soc. of London* 76, pp. 48–61. DOI: 10.1098/rstl.1786.0002 (cit. on p. 49).

Graff, P. et al. (2014). "SkyNet: an efficient and robust neural network training tool for machine learning in astronomy." In: *MNRAS* 441.2, pp. 1741–1759. DOI: 10.1093/mnras/stu642 (cit. on pp. 96, 101, 211).

Graham, A. W. (2002). "The 'Photometric Plane' of elliptical galaxies." In: *MNRAS* 334.4, pp. 859–864. DOI: 10.1046/j.1365-8711.2002.05548.x (cit. on p. 157).

Graham, A. W. and S. P. Driver (2005). "A Concise Reference to (Projected) Sérsic R 1/n Quantities, Including Concentration, Profile Slopes, Petrosian Indices, and Kron Magnitudes." In: *Publications of the Astronomical Society of Australia* 22.2, pp. 118–127. DOI: 10.1071/AS05001 (cit. on pp. 68, 72).

Graham, M. L. et al. (2017). "Photometric Redshifts with the LSST: Evaluating Survey Observing Strategies." In: *AJ* 155.1, p. 1. DOI: 10.3847/1538-3881/aa99d4 (cit. on pp. 101, 211).

Grogin, N. A. et al. (2011). "CANDELS: The Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey." In: *ApJS* 197.2, p. 35. DOI: 10.1088/0067-0049/197/2/35 (cit. on p. 96).

Guinan, E. F. et al. (1998). "The Distance to the Large Magellanic Cloud from the Eclipsing Binary HV 2274." In: *ApJ* 509.1, p. L21. DOI: 10.1086/311760 (cit. on p. 56).

Gunn, J. E. et al. (1998). "The Sloan Digital Sky Survey Photometric Camera." In: *AJ* 116.6, p. 3040. DOI: 10.1086/300645 (cit. on p. 105).

Gunn, J. E. et al. (2006). "The 2.5 m Telescope of the Sloan Digital Sky Survey." In: *AJ* 131.4, p. 2332. DOI: 10.1086/500975 (cit. on pp. 63, 105).

Guo, H., I. Zehavi, and Z. Zheng (2012). "A New Method to Correct for Fiber Collisions in Galaxy Two-point Statistics." In: *ApJ* 756.2, p. 127. DOI: 10.1088/0004-637X/756/2/127 (cit. on p. 78).

Guzzo, L. et al. (2014). "The VIMOS Public Extragalactic Redshift Survey (VIPERS) - An unprecedented view of galaxies and large-scale structure at 0.5<z<1.2." In: *A&A* 566, A108. DOI: 10.1051/0004-6361/201321489 (cit. on p. 121).

Gwyn, S. D. J. and F. D. A. Hartwick (1996). "The Redshift Distribution and Luminosity Functions of Galaxies in the Hubble Deep Field." In: *ApJ* 468.2, p. L77. DOI: 10.1086/310237 (cit. on p. 84).

Hand, N. et al. (2015). "First measurement of the cross-correlation of CMB lensing and galaxy lensing." In: *Phys. Rev. D* 91.6, p. 062001. DOI: 10.1103/PhysRevD.91.062001 (cit. on p. 111).

Herrnstein, J. R. et al. (1999). "A geometric distance to the galaxy NGC4258 from orbital motions in a nuclear gas disk." In: *Nature* 400.6744, pp. 539–541. ISSN: 1476-4687. DOI: 10.1038/22972 (cit. on p. 50).

Hewett, P. C., C. B. Foltz, and F. H. Chaffee (1995). "The large bright quasar survey. 6: Quasar catalog and survey parameters." In: *AJ* 109, pp. 1498–1521. DOI: 10.1086/117380 (cit. on p. 116).

Hildebrandt, H., C. Wolf, and N. Benítez (2008). "A blind test of photometric redshifts on ground-based data." In: *A&A* 480.3, pp. 703–714. DOI: 10.1051/0004-6361:20077107 (cit. on pp. 94, 132).

Hildebrandt, H. et al. (2010). "PHAT: PHoto-z Accuracy Testing." In: *A&A* 523, A31. DOI: 10.1051/0004-6361/201014885 (cit. on pp. 88, 96, 123, 132).

Hildebrandt, H. et al. (2016). "RCSLenS: The Red Cluster Sequence Lensing Survey." In: *MNRAS* 463.1, pp. 635–654. DOI: 10.1093/mnras/stw2013 (cit. on p. 181).

Hoecker, A. et al. (2007). "TMVA - Toolkit for Multivariate Data Analysis." In: *arXiv e-prints*. arXiv: physics/0703039 (cit. on pp. 88, 125, 127).

Hogan, R., M. Fairbairn, and N. Seeburn (2015). "GAz: a genetic algorithm for photometric redshift estimation." In: *MNRAS* 449.2, pp. 2040–2046. DOI: 10.1093/mnras/stv430 (cit. on pp. 89, 101).

Hogg, D. W. (1999). "Distance measures in cosmology." In: *arXiv e-prints*. arXiv: astro-ph/9905116 (cit. on p. 52).

Hogg, D. W. et al. (1998). "A Blind Test of Photometric Redshift Prediction." In: *AJ* 115.4, p. 1418. DOI: 10.1086/300277 (cit. on p. 95).

Hogg, D. W. et al. (2001). "A Photometricity and Extinction Monitor at the Apache Point Observatory." In: *AJ* 122.4, p. 2129. DOI: 10.1086/323103 (cit. on p. 63).

Hogg, D. W. et al. (2002). "The K correction." In: *arXiv e-prints*. arXiv: astro-ph/0210394 (cit. on p. 64).

Hoyle, B. (2016). "Measuring photometric redshifts using galaxy images and Deep Neural Networks." In: *Astron. Comput.* 16, pp. 34–40. DOI: 10.1016/j.ascom.2016.03.006 (cit. on p. 89).

Hoyle, B. et al. (2015a). "Anomaly detection for machine learning redshifts applied to SDSS galaxies." In: *MNRAS* 452.4, pp. 4183–4194. DOI: 10.1093/mnras/stv1551 (cit. on p. 98).

Hoyle, B. et al. (2015b). "Data augmentation for machine learning redshifts applied to Sloan Digital Sky Survey galaxies." In: *MNRAS* 450.1, pp. 305–316. DOI: 10.1093/mnras/stv599 (cit. on p. 88).

Hoyle, B. et al. (2017). "Dark Energy Survey Year 1 Results: Redshift distributions of the weak lensing source galaxies." In: *arXiv e-prints*. arXiv: astro-ph/1708.01532 (cit. on p. 112).

Hsieh, B. C. and H. K. C. Yee (2014). "Estimating Luminosities and Stellar Masses of Galaxies Photometrically without Determining Redshifts." In: *ApJ* 792.2, p. 102. DOI: 10.1088/0004-637X/792/2/102 (cit. on pp. 97, 101).

Hsieh, B. C. et al. (2005). "A Photometric Redshift Galaxy Catalog from the Red-Sequence Cluster Survey." In: *ApJS* 158.2, p. 161. DOI: 10.1086/429293 (cit. on p. 85).

Hubble, E. P. (1926). "Extragalactic nebulae." In: *ApJ* 64, pp. 321–369. DOI: 10.1086/143018 (cit. on p. 69).

— (1929a). "A relation between distance and radial velocity among extra-galactic nebulae." In: *Proceedings of the National Academy of Sciences* 15.3, pp. 168–173. DOI: 10.1073/pnas.15.3.168 (cit. on pp. 40, 42, 51).
— (1929b). "A spiral nebula as a stellar system, Messier 31." In: *ApJ* 69. DOI: 10.1086/143167 (cit. on p. 49).
Huff, E. M. and G. J. Graves (2014). "Magnificent Magnification: Exploiting the Other Half of the Lensing Signal." In: *ApJ* 780.2, p. L16. DOI: 10.1088/2041-8205/780/2/L16 (cit. on p. 157).
Humphreys, E. (2011). "Masers." In: *arXiv e-prints*. arXiv: 1104.2306 (cit. on p. 49).
Huterer, D. et al. (2006). "Systematic errors in future weak-lensing surveys: requirements and prospects for self-calibration." In: *MNRAS* 366.1, pp. 101–114. DOI: 10.1111/j.1365-2966.2005.09782.x (cit. on p. 94).
IAU (1964). "Resolution 4 - On the system of astronomical constants." In: *XII General Assembly of the IAU*. General Assembly of the IAU. Hamburg, Germany: IAU, p. 25. URL: https://www.iau.org/static/resolutions/IAU1964_French.pdf (cit. on p. 44).
— (2012). "Resolution B2 - On the re-definition of the astronomical unit of length." In: *XXVIII General Assembly of the IAU*. General Assembly of the IAU. Beijing, China: IAU, p. 6. URL: https://www.iau.org/static/resolutions/IAU2012_English.pdf (cit. on pp. 43, 45).
— (2015). "Resolution B2 - On recommended zero points for the absolute and apparent bolometric magnitude scales." In: *XXIX General Assembly of the IAU*. General Assembly of the IAU. Honolulu, Hawaii, USA: IAU, p. 13. URL: https://www.iau.org/static/resolutions/IAU2015_English.pdf (cit. on p. 46).
Ibata, R. A. et al. (2017). "The Canada-France Imaging Survey: First Results from the u-Band Component." In: *ApJ* 848.2, p. 128. DOI: 10.3847/1538-4357/aa855c (cit. on p. 181).
Ilbert, O. et al. (2004). "Bias in the estimation of global luminosity functions." In: *MNRAS* 351.2, pp. 541–551. DOI: 10.1111/j.1365-2966.2004.07796.x (cit. on p. 239).
Ilbert, O. et al. (2006). "Accurate photometric redshifts for the CFHT legacy survey calibrated using the VIMOS VLT deep survey." In: *A&A* 457.3, p. 16. DOI: 10.1051/0004-6361:20065138 (cit. on pp. 90, 139–141, 227).
Ilbert, O. et al. (2009). "Cosmos Photometric Redshifts with 30-Bands for 2-deg2." In: *ApJ* 690.2, p. 1236. DOI: 10.1088/0004-637X/690/2/1236 (cit. on pp. 90, 112, 114, 140, 227, 247).
Ivezić, Ž. et al. (2008). "LSST: from Science Drivers to Reference Design and Anticipated Data Products." In: *arXiv e-prints*. arXiv: 0805.2366 (cit. on pp. 82, 105, 112, 113, 181, 210).
Izbicki, R., A. B. Lee, and P. E. Freeman (2017). "Photo-z estimation: An example of nonparametric conditional density estimation under selection bias." In: *Annals of Applied Statistics* 11.2, pp. 698–724. DOI: 10.1214/16-AOAS1013 (cit. on pp. 101, 211).
Jackson, J. C. (1972). "A Critique of Rees's Theory of Primordial Gravitational Radiation." In: *MNRAS* 156.1, 1P–5P. DOI: 10.1093/mnras/156.1.1P (cit. on p. 51).
Jerjen, H. (2001). "Luminosity Function of Galaxies." In: *The Encyclopedia of Astronomy and Astrophysics*. Ed. by P. Murdin. IOP Publishing Ltd. ISBN: 978-0-333-75088-9. DOI: 10.1888/0333750888/2604. URL: http://eaa.crcpress.com/0333750888/2604 (cit. on p. 60).
Jiang, L. et al. (2014). "The Sloan Digital Sky Survey Stripe 82 Imaging Data: Depth-optimized Co-adds over 300 deg2 in Five Filters." In: *ApJS* 213.1, p. 12. DOI: 10.1088/0067-0049/213/1/12 (cit. on p. 108).

Johnson, H. L. and W. W. Morgan (1953). "Fundamental stellar photometry for standards of spectral type on the revised system of the Yerkes spectral atlas." In: *ApJ* 117, p. 313. DOI: 10.1086/145697 (cit. on p. 61).

Johnson, S. P. et al. (2013). "satmc: Spectral energy distribution Analysis Through Markov Chains." In: *MNRAS* 436.3, pp. 2535–2549. DOI: 10.1093/mnras/stt1758 (cit. on pp. 91, 101).

Jones, E. and J. Singal (2017). "Analysis of a custom support vector machine for photometric redshift estimation and the inclusion of galaxy shape information." In: *A&A* 600, A113. DOI: 10.1051/0004-6361/201629558 (cit. on pp. 98, 102, 145).

Jouvel, S. et al. (2017). "Photometric redshifts and clustering of emission line galaxies selected jointly by DES and eBOSS." In: *MNRAS* 469.3, pp. 2771–2790. DOI: 10.1093/mnras/stx163 (cit. on pp. 82, 128).

Jurić, M. et al. (2008). "The Milky Way Tomography with SDSS. I. Stellar Number Density Distribution." In: *ApJ* 673.2, p. 864. DOI: 10.1086/523619 (cit. on p. 48).

Kaiser, N. (1987). "Clustering in real space and in redshift space." In: *MNRAS* 227.1, pp. 1–21. DOI: 10.1093/mnras/227.1.1 (cit. on p. 51).

Kapteyn, J. C. (1922). "First Attempt at a Theory of the Arrangement and Motion of the Sidereal System." In: *ApJ* 55, p. 302. DOI: 10.1086/142670 (cit. on p. 41).

Kazin, E. A. et al. (2014). "The WiggleZ Dark Energy Survey: improved distance measurements to z = 1 with reconstruction of the baryonic acoustic feature." In: *MNRAS* 441.4, pp. 3524–3542. DOI: 10.1093/mnras/stu778 (cit. on p. 56).

Kim, E. J. and R. J. Brunner (2017). "Star-galaxy classification using deep convolutional neural networks." In: *MNRAS* 464.4, pp. 4463–4475. DOI: 10.1093/mnras/stw2672 (cit. on pp. 70, 123).

Kinney, A. L. et al. (1996). "Template Ultraviolet to Near-Infrared Spectra of Star-forming Galaxies and Their Application to K-Corrections." In: *ApJ* 467, p. 38. DOI: 10.1086/177583 (cit. on pp. 90, 139, 140).

Kirchhoff, G. (1860). "Ueber die Fraunhofer'schen Linien." In: *Ann. Phys. (Berl.)* 185.1, pp. 148–150. DOI: 10.1002/andp.18601850115 (cit. on p. 73).

Kochanek, C. S. et al. (2012). "AGES: The AGN and Galaxy Evolution Survey." In: *ApJS* 200.1, p. 8. DOI: 10.1088/0067-0049/200/1/8 (cit. on p. 91).

Koo, D. C. (1985). "Optical multicolors - A poor person's Z machine for galaxies." In: *AJ* 90, pp. 418–440. DOI: 10.1086/113748 (cit. on p. 83).

— (1999). "Photometric Redshifts: A Perspective from an Old-Timer on Its Past, Present, and Potential." In: *arXiv e-prints*. arXiv: astro-ph/9907273 (cit. on pp. 81, 85).

Kotulla, R. and U. Fritze (2009). "Impact of subsolar metallicities on photometric redshifts." In: *MNRAS* 393.1, pp. L55–L59. DOI: 10.1111/j.1745-3933.2008.00598.x (cit. on pp. 91, 101).

Kovalevsky, J. and P. K. Seidelmann (2004). *Fundamentals of Astrometry*. Cambridge University Press. 422 pp. ISBN: 978-1-139-45317-2 (cit. on p. 44).

Kron, R. G. (1980). "Photometry of a complete sample of faint galaxies." In: *ApJS* 43, pp. 305–325. DOI: 10.1086/190669 (cit. on p. 68).

Krone-Martins, A., E. E. O. Ishida, and S. R. De Souza (2014). "The first analytical expression to estimate photometric redshifts suggested by a machine." In: *MNRAS* 443.1, pp. L34–L38. DOI: 10.1093/mnrasl/slu067 (cit. on p. 89).

Kuijken, K. et al. (2015). "Gravitational lensing analysis of the Kilo-Degree Survey." In: *MNRAS* 454.4, pp. 3500–3532. DOI: 10.1093/mnras/stv2140 (cit. on p. 82).

Kuiper, G. P. (1938). "The Empirical Mass-Luminosity Relation." In: *ApJ* 88, p. 472. DOI: 10.1086/143999 (cit. on p. 56).

Kurtz, M. J. et al. (2007). "mu-PhotoZ: Photometric Redshifts by Inverting the Tolman Surface Brightness Test." In: *AJ* 134.4, p. 1360. DOI: 10.1086/521430 (cit. on p. 97).

Kuulkers, E. et al. (2003). "Photospheric radius expansion X-ray bursts as standard candles." In: *A&A* 399.2, pp. 663–680. DOI: 10.1051/0004-6361:20021781 (cit. on p. 57).

LIGO Scientific Collaboration and The Virgo Collaboration et al. (2017). "A gravitational-wave standard siren measurement of the Hubble constant." In: *Nature* 551.7678, pp. 85–88. DOI: 10.1038/nature24471 (cit. on p. 57).

Lahav, O. et al. (1996). "Neural computation as a tool for galaxy classification: methods and examples." In: *MNRAS* 283, p. 207. DOI: 10.1093/mnras/283.1.207 (cit. on pp. 69, 123).

Laigle, C. et al. (2016). "The COSMOS2015 Catalog: Exploring the 1<z<6 Universe with Half a Million Galaxies." In: *ApJS* 224.2, p. 24. DOI: 10.3847/0067-0049/224/2/24 (cit. on pp. 64, 112, 133, 234, 239).

Land, K. et al. (2008). "Galaxy Zoo: the large-scale spin statistics of spiral galaxies in the Sloan Digital Sky Survey." In: *MNRAS* 388.4, pp. 1686–1692. DOI: 10.1111/j.1365-2966.2008.13490.x (cit. on p. 108).

Laureijs, R. et al. (2011). "Euclid Definition Study Report." In: *arXiv e-prints*. arXiv: 1110.3193 (cit. on p. 82).

Laurino, O. et al. (2011). "Astroinformatics of galaxies and quasars: a new general method for photometric redshifts estimation." In: *MNRAS* 418.4, pp. 2165–2195. DOI: 10.1111/j.1365-2966.2011.19416.x (cit. on p. 89).

Lawrence, A. et al. (2007). "The UKIRT Infrared Deep Sky Survey (UKIDSS)." In: *MNRAS* 379.4, pp. 1599–1617. DOI: 10.1111/j.1365-2966.2007.12040.x (cit. on pp. 98, 110).

Le Borgne, J.-F. et al. (2003). "STELIB: A library of stellar spectra at R about 2000." In: *A&A* 402, p. 433. DOI: 10.1051/0004-6361:20030243. (Visited on 05/29/2018) (cit. on p. 139).

Le Fèvre, O. et al. (1994). "CFHT MOS/SIS spectrograph performance." In: *A&A* 282, pp. 325–340. ISSN: 0004-6361. URL: http://adsabs.harvard.edu/abs/1994A%26A...282..325L (visited on 03/01/2018) (cit. on p. 121).

Le Fèvre, O. et al. (2003). "Commissioning and performances of the VLT-VIMOS." In: *Instrument Design and Performance for Optical/Infrared Ground-based Telescopes*. Astronomical Telescopes and Instrumentation. Vol. 4841. Proc. SPIE. Waikoloa, Hawaii, USA: International Society for Optics and Photonics, pp. 1670–1682. DOI: 10.1117/12.460959 (cit. on p. 119).

Le Fèvre, O. et al. (2005). "The VIMOS VLT deep survey - First epoch VVDS-deep survey: 11564 spectra with 17.5<i<24, and the redshift distribution over 0<z<5." In: *A&A* 439.3, pp. 845–862. DOI: 10.1051/0004-6361:20041960 (cit. on pp. 90, 119).

Le Fèvre, O. et al. (2013). "The VIMOS VLT Deep Survey final data release: a spectroscopic sample of 35 016 galaxies and AGN out to z ~ 6.7 selected with 17.5 < iAB < 24.75." In: *A&A* 559, p. 21. DOI: 10.1051/0004-6361/201322179 (cit. on p. 119).

Leauthaud, A. et al. (2017). "Lensing is low: cosmology, galaxy formation or new physics?" In: *MNRAS* 467.3, pp. 3024–3047. DOI: 10.1093/mnras/stx258 (cit. on pp. 111, 208, 209).

Leavitt, H. S. and E. C. Pickering (1912). "Periods of 25 Variable Stars in the Small Magellanic Cloud." In: *Havard College Observatory Circular* 173, pp. 1–3. URL: http://adsabs.harvard.edu/abs/1912HarCi.173....1L (cit. on p. 49).

Leistedt, B. and D. W. Hogg (2017). "Data-driven, Interpretable Photometric Redshifts Trained on Heterogeneous and Unrepresentative Data." In: *ApJ* 838.1, p. 5. DOI: 10.3847/1538-4357/aa6332 (cit. on pp. 88, 90, 102, 134, 136, 137, 211).

Lemaître, A. G. (1931). "A Homogeneous Universe of Constant Mass and Increasing Radius accounting for the Radial Velocity of Extra-galactic Nebulae." In: *MNRAS* 91.5, pp. 483–490. DOI: 10.1093/mnras/91.5.483 (cit. on p. 39).

Li, R. et al. (2016). "Measuring subhalo mass in redMaPPer clusters with CFHT Stripe 82 Survey." In: *MNRAS* 458.3, pp. 2573–2583. DOI: 10.1093/mnras/stw494 (cit. on p. 111).

Lilly, S. J. et al. (2007). "zCOSMOS: A Large VLT/VIMOS Redshift Survey Covering 0<z<3 in the COSMOS Field." In: *ApJS* 172.1, p. 70. DOI: 10.1086/516589 (cit. on pp. 120, 239).

Lima, M. et al. (2008). "Estimating the redshift distribution of photometric galaxy samples." In: *MNRAS* 390.1, pp. 118–130. DOI: 10.1111/j.1365-2966.2008.13510.x (cit. on pp. 95, 176).

Lin, H. et al. (2015). "Are long gamma-ray bursts standard candles?" In: *MNRAS* 453.1, pp. 128–132. DOI: 10.1093/mnras/stv1624 (cit. on p. 57).

Lintott, C. J. et al. (2008). "Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey." In: *MNRAS* 389.3, pp. 1179–1189. DOI: 10.1111/j.1365-2966.2008.13689.x (cit. on pp. 108, 151).

Lintott, C. J. et al. (2011). "Galaxy Zoo 1: data release of morphological classifications for nearly 900 000 galaxies." In: *MNRAS* 410.1, pp. 166–178. DOI: 10.1111/j.1365-2966.2010.17432.x (cit. on pp. 54, 97, 108, 115).

Liske, J. et al. (2015). "Galaxy And Mass Assembly (GAMA): end of survey report and data release 2." In: *MNRAS* 452.2, pp. 2087–2126. DOI: 10.1093/mnras/stv1436 (cit. on p. 98).

Liu, X. et al. (2015). "Cosmological constraints from weak lensing peak statistics with Canada-France-Hawaii Telescope Stripe 82 Survey." In: *MNRAS* 450.3, pp. 2888–2902. DOI: 10.1093/mnras/stv784 (cit. on p. 111).

Loh, E. D. and E. J. Spillar (1986). "Photometric redshifts of galaxies." In: *ApJ* 303, pp. 154–161. DOI: 10.1086/164062 (cit. on pp. 83, 84).

Lupton, R. H., J. E. Gunn, and A. S. Szalay (1999). "A Modified Magnitude System that Produces Well-Behaved Magnitudes, Colors, and Errors Even for Low Signal-to-Noise Ratio Measurements." In: *AJ* 118.3, p. 1406. DOI: 10.1086/301004 (cit. on p. 61).

Macchetto, F. (1976). "The International Ultraviolet Explorer (IUE)." In: *Memorie della Societa Astronomica Italiana* 47, p. 431. URL: https://ui.adsabs.harvard.edu/#abs/1976MmSAI..47..431M/abstract (cit. on p. 139).

Madau, P. et al. (1996). "High-redshift galaxies in the Hubble Deep Field: colour selection and star formation history to z about 4." In: *MNRAS* 283.4, p. 1388. DOI: 10.1093/mnras/283.4.1388 (cit. on pp. 81, 83).

Malling, L. R. and S. W. Golomb (1961). "Radar measurements of the planet venus." In: *Journal of the British Institution of Radio Engineers* 22.4, pp. 297–300. DOI: 10.1049/jbire.1961.0121 (cit. on p. 44).

Malz, A. et al. (2018). "Approximating photo-z PDFs for large surveys." In: *arXiv e-prints*. arXiv: 1806.00014 (cit. on pp. 93, 210).

Mandelbaum, R. et al. (2005). "Systematic errors in weak lensing: application to SDSS galaxy-galaxy weak lensing." In: *MNRAS* 361.4, pp. 1287–1322. DOI: 10.1111/j.1365-2966.2005.09282.x (cit. on p. 82).

Mandelbaum, R. et al. (2008). "Precision photometric redshift calibration for galaxy-galaxy weak lensing." In: *MNRAS* 386.2, pp. 781–806. DOI: 10.1111/j.1365-2966.2008.12947.x (cit. on p. 93).

Margoniner, V. E. and D. M. Wittman (2008). "Photometric Redshifts and Signal-to-Noise Ratios." In: *ApJ* 679.1, p. 31. DOI: 10.1086/528365 (cit. on p. 92).

Markevitch, M. et al. (2004). "Direct Constraints on the Dark Matter Self-Interaction Cross Section from the Merging Galaxy Cluster 1E 0657-56." In: *ApJ* 606.2, p. 819. DOI: 10.1086/383178 (cit. on p. 41).

Marston, P. (2004). *Jeremiah Horrocks: Young Genius & First Venus Transit Observer*. University of Central Lancashire. 40 pp. (cit. on p. 43).

Martí, P. et al. (2014). "Precise photometric redshifts with a narrow-band filter set: the PAU survey at the William Herschel Telescope." In: *MNRAS* 442.1, pp. 92–109. DOI: 10.1093/mnras/stu801 (cit. on pp. 63, 105, 111, 113, 114, 132, 228, 229).

Martin, D. C. et al. (2005). "The Galaxy Evolution Explorer: A Space Ultraviolet Survey Mission." In: *ApJ* 619.1, p. L1. DOI: 10.1086/426387 (cit. on pp. 93, 112).

Massarotti, M., A. Iovino, and A. Buzzoni (2001). "A critical appraisal of the SED fitting method to estimate photometric redshifts." In: *A&A* 368.1, pp. 74–85. DOI: 10.1051/0004-6361:20000553 (cit. on pp. 76, 89).

McCracken, H. J. et al. (2012). "UltraVISTA: a new ultra-deep near-infrared survey in COSMOS." In: *A&A* 544, A156. DOI: 10.1051/0004-6361/201219507 (cit. on p. 112).

McGaugh, S. S. et al. (2000). "The Baryonic Tully-Fisher Relation." In: *ApJ* 533.2, p. L99. DOI: 10.1086/312628 (cit. on p. 54).

McMahon, R. G et al. (2001). "The INT wide field imaging survey (WFS)." In: *New Astronomy Reviews*. Instrumentation at the Isaac Newton Group 45.1, pp. 97–104. DOI: 10.1016/S1387-6473(00)00137-8 (cit. on p. 90).

Milgrom, M. (1983). "A modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis." In: *ApJ* 270, pp. 365–370. DOI: 10.1086/161130 (cit. on p. 41).

Miller, L. et al. (2013). "Bayesian galaxy shape measurement for weak lensing surveys - III. Application to the Canada-France-Hawaii Telescope Lensing Survey." In: *MNRAS* 429.4, pp. 2858–2880. DOI: 10.1093/mnras/sts454 (cit. on p. 204).

Miyazaki, S. et al. (2002). "Subaru Prime Focus Camera - Suprime-Cam." In: *PASJ* 54.6, pp. 833–853. DOI: 10.1093/pasj/54.6.833 (cit. on pp. 105, 112).

Mochejska, B. J. et al. (2000). "The DIRECT Project: Influence of Blending on the Cepheid Distance Scale. I. Cepheids in M31." In: *AJ* 120.2, p. 810. DOI: 10.1086/301493 (cit. on p. 49).

Moffat, A. F. J. (1969). "A Theoretical Investigation of Focal Stellar Images in the Photographic Emulsion and Application to Photographic Photometry." In: *A&A* 3, p. 455. URL: http://adsabs.harvard.edu/abs/1969A%26A.....3..455M (cit. on p. 65).

Molino, A. et al. (2014). "The ALHAMBRA Survey: Bayesian photometric redshifts with 23 bands for 3 deg2." In: *MNRAS* 441.4, pp. 2891–2922. DOI: 10.1093/mnras/stu387 (cit. on p. 227).

Moraes, B. et al. (2014). "The CFHT/MegaCam Stripe-82 Survey." In: *Revista Mexicana de Astronomia y Astrofisica Conference Series*. XIV Latin American Regional IAU Meeting. Vol. 44.

Revista Mexicana de Astronomia y Astrofisica, vol. 27. Florianopolis, Brasil, pp. 202–203. URL: http://www.astroscu.unam.mx/rmaa/RMxAC..44/PDF/RMxAC..44_ABSTRACTS.pdf (cit. on pp. 63, 100, 176).

Moustakas, J. and R. C. Jr. Kennicutt (2006). "An Integrated Spectrophotometric Survey of Nearby Star-forming Galaxies." In: *ApJS* 164.1, p. 81. DOI: 10.1086/500971 (cit. on p. 140).

Murakami, H. et al. (2007). "The Infrared Astronomical Mission AKARI." In: *PASJ* 59 (sp2), S369–S376. DOI: 10.1093/pasj/59.sp2.S369 (cit. on p. 140).

Myers, A. D., M. White, and N. M. Ball (2009). "Incorporating photometric redshift probability density information into real-space clustering measurements." In: *MNRAS* 399.4, pp. 2279–2287. DOI: 10.1111/j.1365-2966.2009.15432.x (cit. on p. 82).

Newcomb, Simon (1906). *A Compendium of Spherical Astronomy with Its Applications to the Determination and Reduction of Positions of the Fixed Stars*. Macmillan Press. 474 pp. (cit. on p. 44).

Newman, J. A. et al. (2013). "The DEEP2 Galaxy Redshift Survey: Design, Observations, Data Reduction, and Redshifts." In: *ApJS* 208.1, p. 5. DOI: 10.1088/0067-0049/208/1/5 (cit. on p. 118).

Niemiec, A. et al. (2017). "Stellar-to-halo mass relation of cluster galaxies." In: *MNRAS* 471.1, pp. 1153–1166. DOI: 10.1093/mnras/stx1667 (cit. on p. 111).

O'Malley, E. M., C. Gilligan, and B. Chaboyer (2017). "Absolute Ages and Distances of 22 GCs Using Monte Carlo Main-sequence Fitting." In: *ApJ* 838.2, p. 162. DOI: 10.3847/1538-4357/aa6574 (cit. on p. 48).

O'Mill, A. L. et al. (2011). "Photometric redshifts and k-corrections for the Sloan Digital Sky Survey Data Release 7." In: *MNRAS* 413.2, pp. 1395–1408. DOI: 10.1111/j.1365-2966.2011.18222.x (cit. on p. 65).

Oke, J. B. and J. E. Gunn (1983). "Secondary standard stars for absolute spectrophotometry." In: *ApJ* 266, pp. 713–717. DOI: 10.1086/160817 (cit. on p. 61).

Ostro, S. J. (1985). "Radar Observations of Asteroids and Comets." In: *PASP* 97.596, p. 877. DOI: 10.1086/131619 (cit. on p. 44).

Oyaizu, H. et al. (2008a). "A Galaxy Photometric Redshift Catalog for the Sloan Digital Sky Survey Data Release 6." In: *ApJ* 674.2, p. 768. DOI: 10.1086/523666 (cit. on pp. 96, 101).

Oyaizu, H. et al. (2008b). "Photometric Redshift Error Estimators." In: *ApJ* 689.2, p. 709. DOI: 10.1086/592591 (cit. on p. 94).

Partridge, R. B. (1974). "A. Search for Primeval Galaxies at High Redshifts." In: *ApJ* 192, p. 241. DOI: 10.1086/153055 (cit. on pp. 81, 82).

Patterson, J. and A. Gibson (2017). *Deep Learning: A Practitioner's Approach*. O'Reilly Media, Inc. 793 pp. ISBN: 978-1-4919-1421-2 (cit. on p. 87).

Peacock, J. A. (1999). *Cosmological Physics*. Cambridge University Press. 700 pp. ISBN: 978-0-521-42270-3 (cit. on pp. 51, 53).

Pelló, R. et al. (1996). "Identification of a high redshift cluster. in the field of Q2345+007 through deep BRIJK' photometry." In: *A&A* 314, pp. 73–86 (cit. on p. 83).

Pereira, M. E. S. et al. (2018). "Weak-lensing calibration of a stellar mass-based mass proxy for redMaPPer and Voronoi Tessellation clusters in SDSS Stripe 82." In: *MNRAS* 474.1, pp. 1361–1372. DOI: 10.1093/mnras/stx2831 (cit. on p. 111).

Perlmutter, S. et al. (1999). "Measurements of Omega and Lambda from 42 High-Redshift Supernovae." In: *ApJ* 517.2, p. 565. DOI: 10.1086/307221 (cit. on p. 40).

Perryman, M. A. C. et al. (1997). "The HIPPARCOS Catalogue." In: *A&A* 323, pp. L49–L52. URL: http://adsabs.harvard.edu/abs/1997A\%26A...323L..49P (cit. on p. 46).

Peters, C. M. et al. (2015). "Quasar Classification Using Color and Variability." In: *ApJ* 811.2, p. 95. DOI: 10.1088/0004-637X/811/2/95 (cit. on pp. 70, 109).

Petrosian, V. (1976). "Surface brightness and evolution of galaxies." In: *ApJ* 209, pp. L1–L5. DOI: 10.1086/182253 (cit. on p. 67).

Pinsonneault, M. H., D. M. Terndrup, and Y. Yuan (2000). "Main sequence fitting and the Hipparcos open cluster distance scale." In: *Stellar Clusters and Associations: Convection, Rotation, and Dynamos*. Ed. by R. Pallavicini, G. Micela, and S. Sciortino. Vol. 198. Astronomical Society of the Pacific Conference, p. 95. ISBN: 1-58381-025-0. URL: http://adsabs.harvard.edu/abs/2000ASPC..198...95P (cit. on p. 48).

Pogson, N. (1856). "Magnitudes of Thirty-six of the Minor Planets for the First Day of each Month of the Year 1857." In: *MNRAS* 17.1, pp. 12–15. DOI: 10.1093/mnras/17.1.12 (cit. on p. 47).

Polletta, M. et al. (2007). "Spectral Energy Distributions of Hard X-Ray Selected Active Galactic Nuclei in the XMM-Newton Medium Deep Survey." In: *ApJ* 663.1, p. 81. DOI: 10.1086/518113 (cit. on pp. 90, 139, 140, 236).

Polsterer, K. L., A. D'Isanto, and F. Gieseke (2016). "Uncertain Photometric Redshifts." In: *arXiv e-prints*. arXiv: 1608.08016 (cit. on pp. 92, 93, 143).

Prevot, M. L. et al. (1984). "The typical interstellar extinction in the Small Magellanic Cloud." In: *A&A* 132, p. 389. URL: https://ui.adsabs.harvard.edu/#abs/1984A&A...132..389P/abstract (cit. on pp. 140, 236).

Prusti, T. et al. (2016). "The Gaia mission." In: *A&A* 595, A1. DOI: 10.1051/0004-6361/201629272 (cit. on p. 46).

Puschell, J. J., F. N. Owen, and R. A. Laing (1982). "Near-infrared photometry of distant radio galaxies - Spectral flux distributions and redshift estimates." In: *ApJ* 257, pp. L57–L61. DOI: 10.1086/183808 (cit. on pp. 82, 83).

Ramos, B. H. F. et al. (2011). "Evolution of Galaxy Luminosity Function Using Photometric Redshifts." In: *AJ* 142.2, p. 41. DOI: 10.1088/0004-6256/142/2/41 (cit. on p. 82).

Ratra, B. and P. J. E. Peebles (1988). "Cosmological consequences of a rolling homogeneous scalar field." In: *Phys. Rev. D* 37.12, pp. 3406–3427. DOI: 10.1103/PhysRevD.37.3406 (cit. on p. 42).

Rau, M. M. et al. (2015). "Accurate photometric redshift probability density estimation - method comparison and application." In: *MNRAS* 452.4, pp. 3710–3725. DOI: 10.1093/mnras/stv1567 (cit. on p. 89).

Reddick, R. M. et al. (2013). "The Connection between Galaxies and Dark Matter Structures in the Local Universe." In: *ApJ* 771.1, p. 30. DOI: 10.1088/0004-637X/771/1/30 (cit. on p. 112).

Reid, M. J. et al. (2009). "The Megamaser Cosmology Project. I. Very Long Baseline Interferometric Observations of UGC 3789." In: *ApJ* 695.1, p. 287. DOI: 10.1088/0004-637X/695/1/287 (cit. on p. 50).

Reis, R. R. R. et al. (2012). "The Sloan Digital Sky Survey Co-add: A Galaxy Photometric Redshift Catalog." In: *ApJ* 747.1, p. 59. DOI: 10.1088/0004-637X/747/1/59 (cit. on pp. 109, 110, 123, 168–170, 206).

Renzini, A. and E. Daddi (2009). "Wandering in the Redshift Desert." In: *The Messenger* 137, pp. 41–45. ISSN: 0722-6691. URL: http://adsabs.harvard.edu/abs/2009Msngr.137...41R (cit. on p. 77).

Richards, G. T. et al. (2001). "Photometric Redshifts of Quasars." In: *AJ* 122.3, p. 1151. DOI: 10.1086/322132 (cit. on p. 86).

Richards, G. T. et al. (2002). "Spectroscopic Target Selection in the Sloan Digital Sky Survey: The Quasar Sample." In: *AJ* 123.6, p. 2945. DOI: 10.1086/340187 (cit. on pp. 77, 116).

Riess, A. G. et al. (1998). "Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant." In: *AJ* 116.3, p. 1009. DOI: 10.1086/300499 (cit. on p. 40).

Robertson, B. E. et al. (2017). "Large Synoptic Survey Telescope Galaxies Science Roadmap." In: *arXiv e-prints*. arXiv: 1708.01617 (cit. on p. 210).

Robertson, H. P. (1935). "Kinematics and World-Structure." In: *ApJ* 82, p. 284. DOI: 10.1086/143681 (cit. on p. 39).

Rozo, E. et al. (2016). "redMaGiC: selecting luminous red galaxies from the DES Science Verification data." In: *MNRAS* 461.2, pp. 1431–1450. DOI: 10.1093/mnras/stw1281 (cit. on p. 173).

Rubin, V. C., W. K. Ford, and N. Thonnard (1980). "Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 (R=4kpc) to UGC 2885 (R=122kpc)." In: *ApJ* 238, pp. 471–487. DOI: 10.1086/158003 (cit. on p. 41).

Russell, H. N. (1928). "On the determination of dynamical parallaxes." In: *AJ* 38, pp. 89–99. DOI: 10.1086/104823 (cit. on p. 56).

Sachs, R. K. and A. M. Wolfe (1967). "Perturbations of a Cosmological Model and Angular Variations of the Microwave Background." In: *ApJ* 147, p. 73. DOI: 10.1086/148982 (cit. on p. 51).

Sadeh, I., F. B. Abdalla, and O. Lahav (2016). "ANNz2: Photometric Redshift and Probability Distribution Function Estimation using Machine Learning." In: *PASP* 128.968, p. 104502. DOI: 10.1088/1538-3873/128/968/104502 (cit. on pp. 88, 92, 101, 127, 131, 211).

Sako, M. et al. (2008). "The Sloan Digital Sky Survey-II Supernova Survey: Search Algorithm and Follow-up Observations." In: *AJ* 135.1, p. 348. DOI: 10.1088/0004-6256/135/1/348 (cit. on pp. 109, 184).

Samui, S. and S. Samui Pal (2017). "Photo-z with CuBANz: An improved photometric redshift estimator using Clustering aided Back propagation Neural network." In: *New Astronomy* 51, pp. 169–177. DOI: 10.1016/j.newast.2016.09.002 (cit. on p. 101).

Sánchez, C. et al. (2014). "Photometric redshift analysis in the Dark Energy Survey Science Verification data." In: *MNRAS* 445.2, pp. 1482–1506. DOI: 10.1093/mnras/stu1836 (cit. on pp. 89, 95, 96, 123, 128, 176, 177).

Sandage, A., R. L. Beaton, and S. R. Majewski (2016). "Comparison of Hipparcos Trigonometric and Mount Wilson Spectroscopic Parallaxes for 90 Subgiants that Defined the Class in 1935." In: *PASP* 128.964, p. 064202. DOI: 10.1088/1538-3873/128/964/064202 (cit. on p. 48).

Saulder, C. et al. (2013). "Calibrating the fundamental plane with SDSS DR8 data." In: *A&A* 557, A21. DOI: 10.1051/0004-6361/201321466 (cit. on pp. 54, 65, 108, 115, 154, 156, 157).

Sawicki, M. J., H. Lin, and H. K. C. Yee (1997). "Evolution of the Galaxy Population Based on Photometric Redshifts in the Hubble Deep Field." In: *AJ* 113, p. 1. DOI: 10.1086/118231 (cit. on p. 84).

Schechter, P. (1976). "An analytic expression for the luminosity function for galaxies." In: *ApJ* 203, pp. 297–306. DOI: 10.1086/154079 (cit. on pp. 59, 237).

Schlafly, E. F. and D. P. Finkbeiner (2011). "Measuring Reddening with Sloan Digital Sky Survey Stellar Spectra and Recalibrating SFD." In: *ApJ* 737.2, p. 103. DOI: 10.1088/0004-637X/737/2/103 (cit. on p. 64).

Schlegel, D. J., D. P. Finkbeiner, and M. Davis (1998). "Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds." In: *ApJ* 500.2, p. 525. DOI: 10.1086/305772 (cit. on p. 64).

Schlegel, D. J. et al. (2015). "The Dark Energy Spectroscopic Instrument (DESI): The NOAO DECam Legacy Imaging Survey and DESI Target Selection." In: *225th Meetng of the American Astronomical Society*. American Astronomical Society Meeting Abstracts. Seattle, WA, US, p. 336.07. URL: https://files.aas.org/aas225/AAS_225_Abstracts.pdf (cit. on pp. 108, 181).

Schmidt, M. (1968). "Space Distribution and Luminosity Functions of Quasi-Stellar Radio Sources." In: *ApJ* 151, p. 393. DOI: 10.1086/149446 (cit. on p. 237).

Schneider, G., J. M. Pasachoff, and R. C. Willson (2006). "The Effect of the Transit of Venus on ACRIM's Total Solar Irradiance Measurements: Implications for Transit Studies of Extrasolar Planets." In: *ApJ* 641.1, p. 565. DOI: 10.1086/500427 (cit. on p. 43).

Scoville, N. et al. (2007). "The Cosmic Evolution Survey (COSMOS): Overview." In: *ApJS* 172.1, p. 1. DOI: 10.1086/516585 (cit. on p. 112).

Scowcroft, V. et al. (2009). "The effect of metallicity on Cepheid magnitudes and the distance to M33." In: *MNRAS* 396.3, pp. 1287–1296. DOI: 10.1111/j.1365-2966.2009.14822.x (cit. on p. 49).

Sérsic, J. L. (1963). "Influence of the atmospheric and instrumental dispersion on the brightness distribution in a galaxy." In: *Boletin de la Asociacion Argentina de Astronomia La Plata Argentina* 6, p. 41. URL: http://adsabs.harvard.edu/abs/1963BAAA....6...41S (cit. on p. 66).

Shan, H. et al. (2014). "Weak lensing mass map and peak statistics in Canada-France-Hawaii Telescope Stripe 82 survey." In: *MNRAS* 442.3, pp. 2534–2542. DOI: 10.1093/mnras/stu1040 (cit. on p. 111).

Shan, H. et al. (2017). "The Mass-Concentration Relation and the Stellar-to-halo Mass Ratio in the CFHT Stripe 82 Survey." In: *ApJ* 840.2, p. 104. DOI: 10.3847/1538-4357/aa6c68 (cit. on p. 111).

Sharp, R. et al. (2006). "Performance of AAOmega: the AAT multi-purpose fiber-fed spectrograph." In: *Ground-based and Airborne Instrumentation for Astronomy*. SPIE Astronomical Telescopes + Instrumentation. Vol. 6269. Proc. SPIE. Orlando, Florida, USA: International Society for Optics and Photonics, 62690G. DOI: 10.1117/12.671022 (cit. on p. 119).

Silva, L. et al. (1998). "Modeling the Effects of Dust on Galactic Spectral Energy Distributions from the Ultraviolet to the Millimeter Band." In: *ApJ* 509.1, p. 103. DOI: 10.1086/306476 (cit. on p. 140).

Singal, J. et al. (2011). "The Efficacy of Galaxy Shape Parameters in Photometric Redshift Estimation: A Neural Network Approach." In: *PASP* 123.903, pp. 615–621. DOI: 10.1086/660155 (cit. on p. 97).

Skrutskie, M. F. et al. (2006). "The Two Micron All Sky Survey (2MASS)." In: *AJ* 131.2, p. 1163. DOI: 10.1086/498708 (cit. on p. 110).

Smee, S. A. et al. (2013). "The Multi-object, Fiber-fed Spectrographs for the Sloan Digital Sky Survey and the Baryon Oscillation Spectroscopic Survey." In: *AJ* 146.2, p. 32. DOI: 10.1088/0004-6256/146/2/32 (cit. on pp. 115, 117).

Smith, J. A. et al. (2002). "The ugriz Standard-Star System." In: *AJ* 123.4, p. 2121. ISSN: 1538-3881. DOI: 10.1086/339311. URL: http://stacks.iop.org/1538-3881/123/i=4/a=2121 (visited on 02/22/2018) (cit. on p. 61).

Smith, W. (1769). "The Sun's Parallax Deduced from a Comparison of the Norriton Observations of the Transit of Venus, 1769; With the Greenwich and Other European Observations of the Same." In: *Transactions of the American Philosophical Society* 1, pp. 54–72. ISSN: 0065-9746. DOI: 10.2307/1005019 (cit. on p. 43).

Smullin, L. D. and G. Fiocco (1962). "Optical Echoes from the Moon." In: *Nature* 194.4835, p. 1267. DOI: 10.1038/1941267a0 (cit. on p. 44).

Soltan, A. M. and M. J. Chodorowski (2015). "Galaxy clustering using photometric redshifts." In: *MNRAS* 453.1, pp. 1013–1025. DOI: 10.1093/mnras/stv1664 (cit. on p. 82).

Soo, J. Y. H. et al. (2018). "Morpho-z: improving photometric redshifts with galaxy morphology." In: *MNRAS* 475.3, pp. 3613–3632. DOI: 10.1093/mnras/stx3201 (cit. on pp. 100, 123, 146, 191).

Soumagnac, M. T. et al. (2015). "Star/galaxy separation at faint magnitudes: application to a simulated Dark Energy Survey." In: *MNRAS* 450.1, pp. 666–680. DOI: 10.1093/mnras/stu1410 (cit. on pp. 70, 123, 168).

Sowards-Emmerd, D. et al. (2000). "A Catalog of Photometry for Las Campanas Redshift Survey Galaxies on the Sloan Digital Sky Survey System." In: *AJ* 119.6, p. 2598. DOI: 10.1086/301394 (cit. on p. 85).

Spergel, D. et al. (2013). "Wide-Field InfraRed Survey Telescope-Astrophysics Focused Telescope Assets WFIRST-AFTA Final Report." In: *arXiv e-prints*. arXiv: 1305.5422. URL: http://arxiv.org/abs/1305.5422 (visited on 04/21/2018) (cit. on p. 82).

Springel, V. (2005). "The cosmological simulation code gadget-2." In: *MNRAS* 364.4, pp. 1105–1134. DOI: 10.1111/j.1365-2966.2005.09655.x (cit. on p. 114).

Stabenau, H. F., A. Connolly, and B. Jain (2008). "Photometric redshifts with surface brightness priors." In: *MNRAS* 387.3, pp. 1215–1226. DOI: 10.1111/j.1365-2966.2008.13317.x (cit. on p. 97).

Steidel, C. C. et al. (1996). "Spectroscopy of Lyman Break Galaxies in the Hubble Deep Field." In: *AJ* 112, p. 352. DOI: 10.1086/118019 (cit. on pp. 82, 83).

Steidel, C. C. et al. (2004). "A Survey of Star-forming Galaxies in the 1.4<z<2.5 Redshift Desert: Overview." In: *ApJ* 604.2, p. 534. DOI: 10.1086/381960 (cit. on p. 120).

Steinhardt, C. L. et al. (2014). "Star Formation at 4<z<6 from the Spitzer Large Area Survey with Hyper-Suprime-Cam (SPLASH)." In: *ApJ* 791.2, p. L25. DOI: 10.1088/2041-8205/791/2/L25 (cit. on p. 112).

Stimson, G. W. (1998). *Introduction to Airborne Radar*. SciTech Pub. 600 pp. ISBN: 978-1-891121-01-2 (cit. on p. 44).

Stoughton, C. et al. (2002). "Sloan Digital Sky Survey: Early Data Release." In: *AJ* 123.1, p. 485. DOI: 10.1086/324741 (cit. on pp. 67, 70).

Strateva, I. et al. (2001). "Color Separation of Galaxy Types in the Sloan Digital Sky Survey Imaging Data." In: *AJ* 122.4, p. 1861. DOI: 10.1086/323301 (cit. on p. 72).

Strauss, M. A. et al. (2002). "Spectroscopic Target Selection in the Sloan Digital Sky Survey: The Main Galaxy Sample." In: *AJ* 124.3, p. 1810. DOI: 10.1086/342343 (cit. on pp. 67, 108, 115).

Strelnitski, V. (1997). "Masers, Lasers and the Interstellar Medium." In: *Astrophysics and Space Science* 252.1, pp. 279–287. DOI: 10.1023/A:1000892300429 (cit. on p. 50).

SubbaRao, M. U. et al. (1996). "Luminosity Functions From Photometric Redshifts. I. Techniques." In: *AJ* 112, p. 929. DOI: 10.1086/118066 (cit. on pp. 82, 85, 91).

Tagliaferri, R. et al. (2003). "Neural Networks for Photometric Redshifts Evaluation." In: *WIRN 2003: Neural Nets*. Italian Workshop on Neural Nets. Vol. 2859. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 226–234. ISBN: 978-3-540-45216-4. DOI: 10.1007/978-3-540-45216-4_26 (cit. on p. 97).

Tanaka, M. (2015). "Photometric Redshift with Bayesian Priors on Physical Properties of Galaxies." In: *ApJ* 801.1, p. 20. DOI: 10.1088/0004-637X/801/1/20 (cit. on pp. 97, 101).

Tanaka, M. et al. (2018). "Photometric redshifts for Hyper Suprime-Cam Subaru Strategic Program Data Release 1." In: *PASJ* 70 (SP1). DOI: 10.1093/pasj/psx077 (cit. on pp. 93, 97, 101).

Teets, D. A. (2003). "Transits of Venus and the Astronomical Unit." In: *Mathematics Magazine* 76.5, pp. 335–348. URL: http://www.jstor.org/stable/3654879 (cit. on p. 44).

Terlevich, R. et al. (1981). "The metallicities, velocity dispersions and true shapes of elliptical galaxies." In: *MNRAS* 196.2, pp. 381–395. DOI: 10.1093/mnras/196.2.381 (cit. on p. 53).

Thakar, A. R. et al. (2008). "The Catalog Archive Server Database Management System." In: *Computing in Science Engineering* 10.1, pp. 30–37. DOI: 10.1109/MCSE.2008.15 (cit. on p. 108).

Thomas, S. A., F. B. Abdalla, and O. Lahav (2010). "Upper Bound of 0.28 eV on Neutrino Masses from the Largest Photometric Redshift Survey." In: *Physics Review Letters* 105.3, p. 031301. DOI: 10.1103/PhysRevLett.105.031301 (cit. on p. 82).

Thuan, T. X. and J. E. Gunn (1976). "A New Four-Color Intermediate-band Photometric System." In: *PASP* 88.524, p. 543. DOI: 10.1086/129982 (cit. on p. 61).

Tolman, R. C. (1930). "On the Estimation of Distances in a Curved Universe with a Non-Static Line Element." In: *Proceedings of the National Academy of Sciences of the United States of America* 16.7, pp. 511–520. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC526678/ (cit. on p. 72).

Tonry, J. and M. Davis (1979). "A survey of galaxy redshifts. I - Data reduction techniques." In: *AJ* 84, pp. 1511–1525. DOI: 10.1086/112569 (cit. on p. 78).

Tully, R. B. and J. R. Fisher (1977). "A new method of determining distances to galaxies." In: *A&A* 54, pp. 661–673. URL: http://adsabs.harvard.edu/abs/1977A%26A....54..661T (cit. on p. 54).

Turner, D. G. (1996). "The Progenitors of Classical Cepheid Variables." In: *Journal of the Royal Astronomical Society of Canada* 90, p. 82 (cit. on p. 48).

Van Helden, A. (2010). *Measuring the Universe: Cosmic Dimensions from Aristarchus to Halley*. University of Chicago Press. 213 pp. ISBN: 978-0-226-84890-7 (cit. on p. 43).

Vanzella, E. et al. (2004). "Photometric redshifts with the Multilayer Perceptron Neural Network: Application to the HDF-S and SDSS." In: *A&A* 423.2, p. 16. DOI: 10.1051/0004-6361:20040176 (cit. on p. 88).

Vikhlinin, A. et al. (2006). "Chandra Sample of Nearby Relaxed Galaxy Clusters: Mass, Gas Fraction, and Mass-Temperature Relation." In: *ApJ* 640.2, p. 691. DOI: 10.1086/500288 (cit. on p. 41).

Vince, O. and I. Csabai (2006). "Toward more precise photometric redshift estimation." In: *Proc. IAU* 2 (S241), pp. 573–574. DOI: 10.1017/S1743921307009179 (cit. on p. 97).

Vincent, R. A. and B. S. Ryden (2005). "The Dependence of Galaxy Shape on Luminosity and Surface Brightness Profile." In: *ApJ* 623.1, p. 137. DOI: 10.1086/428765 (cit. on p. 72).

Vitorelli, A. Z. et al. (2018). "On mass concentrations and magnitude gaps of galaxy systems in the CS82 survey." In: *MNRAS* 474.1, pp. 866–875. DOI: 10.1093/mnras/stx2791 (cit. on p. 111).

Wadadekar, Y. (2005). "Estimating Photometric Redshifts Using Support Vector Machines." In: *PASP* 117.827, pp. 79–85. DOI: 10.1086/427710 (cit. on pp. 88, 89, 97).

Wake, D. A., P. G. Van Dokkum, and M. Franx (2012). "Revealing Velocity Dispersion as the Best Indicator of a Galaxy's Color, Compared to Stellar Mass, Surface Mass Density, or Morphology." In: *ApJ* 751.2, p. L44. DOI: 10.1088/2041-8205/751/2/L44 (cit. on p. 157).

Walker, A. G. (1937). "On Milne's Theory of World-Structure." In: *Proceedings of the London Mathematical Society* s2-42.1, pp. 90–127. DOI: 10.1112/plms/s2-42.1.90 (cit. on p. 39).

Wang, D. et al. (2007). "Kernel regression for determining photometric redshifts from Sloan broad-band photometry." In: *MNRAS* 382.4, pp. 1601–1606. DOI: 10.1111/j.1365-2966.2007.12129.x (cit. on p. 89).

Wang, D., Y. Zhang, and Y. Zhao (2008a). "Predicting photometric redshifts with polynomial regression." In: *Advanced Software and Control for Astronomy II*. SPIE Astronomical Telescopes + Instrumentation. Vol. 7019. Proc. SPIE. Marseille, France: International Society for Optics and Photonics, 70193A. DOI: 10.1117/12.788443 (cit. on p. 86).

Wang, D. et al. (2008b). "Two Novel Approaches for Photometric Redshift Estimation based on SDSS and 2MASS." In: *ChJAA* 8.1, p. 119. DOI: 10.1088/1009-9271/8/1/13 (cit. on pp. 88, 95).

Wang, D., Y. Zhang, and Y. Zhao (2010). "An automated algorithm for determining photometric redshifts of quasars." In: *Software and Cyberinfrastructure for Astronomy*. SPIE Astronomical Telescopes + Instrumentation. Vol. 7740. Proc. SPIE. San Diego, California, USA: International Society for Optics and Photonics, 77402Y. DOI: 10.1117/12.856815 (cit. on p. 97).

Way, M. J. (2011). "Galaxy Zoo Morphology and Photometric Redshifts in the Sloan Digital Sky Survey." In: *ApJ* 734.1, p. L9. DOI: 10.1088/2041-8205/734/1/L9 (cit. on pp. 97, 108, 145, 151, 154).

Way, M. J. and C. D. Klose (2012). "Can Self-Organizing Maps Accurately Predict Photometric Redshifts?" In: *PASP* 124.913, p. 274. DOI: 10.1086/664796 (cit. on p. 89).

Way, M. J. and A. N. Srivastava (2006). "Novel methods for predicting photometric redshifts from broadband photometry using virtual sensors." In: *ApJ* 647.1, p. 102. DOI: 10.1086/505293 (cit. on pp. 71, 72, 89).

Way, M. J. et al. (2009). "New Approaches to Photometric Redshift Prediction via Gaussian Process Regression in the Sloan Digital Sky Survey." In: *ApJ* 706.1, pp. 623–636. DOI: 10.1088/0004-637X/706/1/623 (cit. on p. 97).

Weaver, H. et al. (1965). "Observations of a Strong Unidentified Microwave Line and of Emission from the OH Molecule." In: *Nature* 208.5005, pp. 29–31. DOI: 10.1038/208029a0 (cit. on p. 50).

Weinstein, M. A. et al. (2004). "An Empirical Algorithm for Broadband Photometric Redshifts of Quasars from the Sloan Digital Sky Survey." In: *ApJS* 155.2, p. 243. DOI: 10.1086/425355 (cit. on p. 86).

Werner, M. W. et al. (2004). "The Spitzer Space Telescope Mission." In: *ApJS* 154.1, p. 1. DOI: 10.1086/422992 (cit. on p. 140).

Wesselius, P. R. et al. (1982). "ANS ultraviolet photometry, catalogue of point sources." In: *A&AS* 49, p. 427. URL: https://ui.adsabs.harvard.edu/#abs/1982A&AS...49..427W/abstract (cit. on p. 139).

Wiklind, T. et al. (2008). "A Population of Massive and Evolved Galaxies at z>5." In: *ApJ* 676.2, p. 781. DOI: 10.1086/524919 (cit. on p. 101).

Williams, R. E. et al. (1996). "The Hubble Deep Field: Observations, Data Reduction, and Galaxy Photometry." In: *AJ* 112, p. 1335. DOI: 10.1086/118105 (cit. on p. 133).

Willmer, C. N. A. (1997). "Estimating galaxy luminosity functions." In: *AJ* 114, pp. 898–912. DOI: 10.1086/118522 (cit. on p. 239).

Wittman, D. M. (2009). "What Lies Beneath: Using p(z) to Reduce Systematic Photometric Redshift Errors." In: *ApJ* 700.2, p. L174. DOI: 10.1088/0004-637X/700/2/L174 (cit. on p. 93).

Wittman, D. M., R. Bhaskar, and R. Tobin (2016). "Overconfidence in photometric redshift estimation." In: *MNRAS* 457.4, pp. 4005–4011. DOI: 10.1093/mnras/stw261 (cit. on p. 93).

Wolf, C. (2009). "Bayesian photometric redshifts with empirical training sets." In: *MNRAS* 397.1, pp. 520–533. DOI: 10.1111/j.1365-2966.2009.14953.x (cit. on pp. 90, 93, 96, 102).

Wollaston, W. H. (1802). "XII. A method of examining refractive and dispersive powers, by prismatic reflection." In: *Phil. Trans. R. Soc. of London* 92, pp. 365–380. DOI: 10.1098/rstl.1802.0014 (cit. on p. 73).

Wray, J. J. and J. E. Gunn (2008). "A New Technique for Galaxy Photometric Redshifts in the Sloan Digital Sky Survey." In: *ApJ* 678.1, pp. 144–153. DOI: 10.1086/529127 (cit. on p. 97).

Wright, E. L. et al. (2010). "The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance." In: *AJ* 140.6, p. 1868. DOI: 10.1088/0004-6256/140/6/1868 (cit. on p. 227).

Xia, L. et al. (2009). "Improved Photometric Redshifts with Surface Luminosity Priors." In: *AJ* 138.1, p. 95. DOI: 10.1088/0004-6256/138/1/95 (cit. on p. 97).

Yee, H. K. C. et al. (2000). "The CNOC2 Field Galaxy Redshift Survey. I. The Survey and the Catalog for the Patch CNOC 0223+00." In: *ApJS* 129.2, p. 475. DOI: 10.1086/313426 (cit. on p. 121).

York, D. G. et al. (2000). "The Sloan Digital Sky Survey: Technical Summary." In: *AJ* 120.3, p. 1579. DOI: 10.1086/301513 (cit. on pp. 105, 109, 115, 184).

Zehavi, I. et al. (2002). "Galaxy Clustering in Early Sloan Digital Sky Survey Redshift Data." In: *ApJ* 571.1, p. 172. DOI: 10.1086/339893 (cit. on p. 115).

Zheng, H. and Y. Zhang (2012). "Review of techniques for photometric redshift estimation." In: *Software and Cyberinfrastructure for Astronomy II*. SPIE Astronomical Telescopes + Instrumentation. Vol. 8451. Proc. SPIE. Amsterdam, Netherlands: International Society for Optics and Photonics, p. 845134. DOI: 10.1117/12.925314 (cit. on p. 90).

Zitlau, R. et al. (2016). "Stacking for machine learning redshifts applied to SDSS galaxies." In: *MNRAS* 460.3, pp. 3152–3162. DOI: 10.1093/mnras/stw1454 (cit. on p. 98).

Zou, H. et al. (2017). "Project Overview of the Beijing-Arizona Sky Survey." In: *PASP* 129.976, p. 064101. DOI: 10.1088/1538-3873/aa65ba (cit. on p. 181).

Zwicky, F. (1933). "Die Rotverschiebung von extragalaktischen Nebeln." In: *Helvetica Physica Acta* 6, pp. 110–127. URL: http://adsabs.harvard.edu/abs/1933AcHPh...6..110Z (cit. on p. 41).

de Jong, J. T. A. et al. (2013). "The Kilo-Degree Survey." In: *Exp. Astron.* 35.1, pp. 25–44. DOI: 10.1007/s10686-012-9306-1 (cit. on pp. 82, 181).

de Sitter, W. (1911). "On the bearing of the Principle of Relativity on Gravitational Astronomy." In: *MNRAS* 71.5, pp. 388–415. DOI: 10.1093/mnras/71.5.388 (cit. on p. 39).

de Vaucouleurs, G. (1948). "Recherches sur les Nebuleuses Extragalactiques." In: *Annales d'Astrophysique* 11, p. 247. URL: http://adsabs.harvard.edu/abs/1948AnAp...11..247D (cit. on p. 66).

van de Hulst, H. C., C. A. Muller, and J. H. Oort (1954). "The spiral structure of the outer part of the Galactic System derived from the hydrogen emission at 21 cm wavelength." In: *Bulletin of the Astronomical Institutes of the Netherlands* 12, p. 117. URL: http://adsabs.harvard.edu/abs/1954BAN....12..117V (cit. on p. 75).